DOCUMENT RESUME

ED 050 228

VT 011 130

AUTHOR	Gonzalez, Maria Elena
TITLE	An Optimal Sample Design for a Job Vacancy Survey.
INSTITUTION	National Industrial Conference Board, Inc., New York, N.Y.
SPONS AGENCY	Office of Manpower Policy, Evaluation, and Research (DOL), Washington, D.C.
PUB DATE	Jun 68
NOTE	129p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS	Employment Opportunities, *Labor Economics, Labor Force, Labor Market, *Manpower Needs, *Measurement Techniques, Occupational Surveys, *Research Methodology

ABSTRACT

Job vacancy statistics are useful both nationally and locally to manpower policymakers in: (1) measuring labor shortages or surpluses, (2) identifying employment trends in occupations and industries, and (3) matching persons and jobs. The purposes of this study are to analyze the reliability of vacancy estimates and to plan efficient sample designs of job vacancy surveys. The report describes methods of reliability measurement for estimates of job vacancies at a point in time and for estimates of changes between surveys. The study found that sample design efficiency could be improved significantly by: (1) stratification by size of firm and by industry, (2) use of smaller employers in samples, and (3) measurement of changes in job vacancies, rather than total vacancies. A related report is available as ED 043 726. (BH)

ED050228

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

AN OPTIMAL SAMPLE DESIGN FOR A JOB VACANCY SURVEY

by

María Elena González

Prepared by

National Industrial Conference Board

for

The Office of Manpower Policy, Evaluation, and Research
Manpower Administration
U. S. Department of Labor

June, 1968

VT011130

This report was prepared under Contract No. 81-34-66-13 for the Manpower Administration, U.S. Department of Labor, under the authority of the Manpower Development and Training Act. Researchers undertaking such projects under the Government sponsorship are encouraged to express their own judgment. Interpretations or viewpoints stated in this document do not necessarily represent the official position or policy of the Department of Labor.

TABLE OF CONTENTS

Page

LIST OF TABLES v

ACKNOWLEDGMENTS viii

I- INTRODUCTION AND SUMMARY . . . . I: 1-11

Criteria for Sample Adequacy . . . . .I - 3

Summary of Report . . . . I - 6

II- STRATIFICATION AND DATA SOURCE . . . . II: 1-9

A. Stratification in a Job Vacancy Survey . . . . II - 1

B. Selection of the NICB Sample . . . . II - 4

Summary . . . . .II - 8

III- MEASURES OF VARIABILITY . . . . .III: 1-29

A. Actual Measures of Variability for Each Survey Period . .III - 3

1. Vacancies by industry . . . . .III - 3

2. Vacancies by size . . . . .III - 8

3. Vacancy rate by industry . . . . III - 11

4. Vacancy rate by size . . . . III - 14

B. Measures of Variability for Changes Between Survey Periods . . . . .III - 16

1. Changes in vacancies by industry . . . . III - 16

2. Changes in vacancies by size . . . . III - 19

3. Changes in vacancy rate by industry . . . . .III - 22

4. Changes in vacancy rate by size . . . . .III - 26

Summary . . . . .III - 28

IV- A PROPOSED OPTIMAL SAMPLE DESIGN . . . . IV: 1-44

A. Selection of Strata for Optimal Allocation . . . . IV - 1

B. Methods for Obtaining an Optimal Sample Design . . . . . IV - 2

Page

C. Optimal Sample Design for Estimating Total Number of Job Vacancies . . . . IV - 13

D. Optimal Sample Design for Estimating Changes in Number of Total Job Vacancies . . . . IV - 19

E. Optimal Sample Design for Estimating the Job Vacancy Rate . . . . IV - 24

F. Optimal Sample Design for Estimating Changes in the Job Vacancy Rate . . . . IV - 29

G. Comparison of Sample Designs . . . . IV - 35

Summary . . . . IV - 43

V- GENERAL METHODS FOR DETERMINING STRATA AND SAMPLE SIZES . V: 1 - 14

A. The Choice of Stratification Variables . . . . V - 1

B. The Choice of the Number of Strata . . . . V - 2

C. The Formation of Strata in the Population . . . . . . V - 3

D. n_g: The Sample Size in Stratum g . . . . V - 5

Summary . . . . V - 13

APPENDIX A - PROOFS OF UNBIASED AND CONSISTENT ESTIMATES OF ELEMENT VARIANCE AND ELEMENT COVARIANCE . A: 1 - 13

BIBLIOGRAPHY . . . . B - 1

LIST OF TABLES

Page

1. Total Number of Job Vacancies, Standard Error, and Coefficient of Variation, by Industry Group III-7

2. Total Number of Job Vacancies, Standard Error, and Coefficient of Variation, by Employment Size III-10

3. Job Vacancy Rate, Standard Error, and Coefficient of Variation, by Industry Group III-13

4. Job Vacancy Rate, Standard Error, and Coefficient of Variation, by Employment Size III-15

5. Changes in Job Vacancies, Standard Error, and Coefficient of Variation, by Industry Group III-18

6. Changes in Job Vacancies, Standard Error, and Coefficient of Variation, by Employment Size III-20

7. Changes in the Job Vacancy Rate, Standard Error, and Coefficient of Variation, by Industry Group III-24

8. Changes in the Job Vacancy Rate, Standard Error, and Coefficient of Variation, by Employment Size III-27

9. Element Standard Error for Total Job Vacancies and for Changes in Total Job Vacancies, and Estimated Total Survey Cost Per Firm IV-6

10. Element Standard Error for the Total Job Vacancy Rate and for Changes in the Total Job Vacancy Rate IV-10

11. Optimal Sample Size for Estimates of Total Job Vacancies . . . IV-14

12. Optimal Sampling Ratios of Firms for Estimates of Total Job Vacancies IV-15

13. Average Optimal Sample Size and Largest Optimal Sample Size for Total Job Vacancies IV-17

14. Selected Sample Design for Estimating Total Job Vacancies to Obtain a Coefficient of Variation of 0.05 . . . . IV-18

15. Optimal Sample Size for Estimates of Changes in Total Job Vacancies IV-20

16. Selected Sample Design for Estimating Changes in Total Job Vacancies Which Measure 0.5% of (E+V) with 95% Confidence . . IV-21

Page

17. Optimal Sampling Ratios of Firms for Estimates of Changes in Total Job Vacancies IV-23

18. Largest Optimal Sample Size for Each Stratum to Estimate Three-month Changes in Vacancies of 0.5% of (E+V) with 95% Confidence IV-24

19. Optimal Sample Size for Estimates of the Total Job Vacancy Rate IV-26

20. Optimal Sampling Ratios of Firms for Estimates of the Total Job Vacancy Rate IV-27

21. Selected Sample Design for Estimating the Job Vacancy Rate with a Coefficient of Variation of 0.05 . . . . IV-28

22. Largest Optimal Sample Size for Each Stratum to Estimate the Job Vacancy Rate with a Coefficient of Variation of 0.05. . . IV-29

23. Optimal Sample Size for Estimates of Changes in the Total Job Vacancy Rate IV-30

24. Selected Sample Design for Estimating Three-month Changes of 0.5 or More in the Job Vacancy Rate with 95% Confidence. . . IV-31

25. Optimal Sampling Ratios of Firms for Estimates of Changes in the Total Job Vacancy Rate IV-32

26. Average Optimal Sample Size for Vacancy Rates and Three-month Changes in Vacancy Rates IV-33

27. Largest Optimal Sample Size of Each Stratum for Estimating Three-month Changes of 0.5 or More in the Job Vacancy Rate with 95% Confidence IV-34

28. Per Cent Distribution of Employer Units in Optimal Sample Designs for Monroe County Compared with the NICB Surveys of Monroe County IV-36

29. Sample Size and Cost of Job Vacancy Surveys IV-38

30. Strata in Which Estimates for Changes Exceed Those for Totals: Percentages of Total Samples and Their Ratios IV-39

31. Ratio of Estimates for One Survey Period to Estimates for Three-month Changes for: Optimal Sample Size and Cost . . . IV-40

32. Largest Sample Size for Each Stratum of the Designs Given in Sections C to F IV-41

Page

33. Strata in Which Optimal Percentages Are Larger for Estimates of Vacancy Rates than for Estimates of Job Vacancies IV-42

34. Largest Sample Size for Each Stratum of the Designs to Estimate Three-month Changes Given in Tables 18 and 27 . . . . . IV-43

35. Optimal Size Strata for Rochester, N.Y., New York, N.Y., and Richmond, Virginia, in 1965 . . . . V-4

36. Estimated Rate of Current Job Vacancies by Area, April 1, 1966 and April 15, 1965 . . . . V-5

37. Element Standard Errors for Total Job Vacancies and Estimated Survey Cost Per Firm for Rochester Surveys in 1965 . . . . . . V-8

38. Optimal Sample Sizes for Estimating Total Job Vacancies in Rochester, 1965, and for a Job Vacancy Rate of 1.5 . . . . . . V-10

39. Largest Optimal Sample Size of Each Stratum for Estimating Total Job Vacancies Found in Rochester with Coefficients of Variation of 0.05 and 0.10 . . . . V-11

40. Average Optimal Sample Sizes for Estimating Total Job Vacancies According to Specified Job Vacancy Rates V-12

ACKNOWLEDGMENTS

The present study is one of several growing out of NICB's earlier report, Measuring Job Vacancies. This series of studies has been made possible by a generous grant from the Office of Manpower Policy, Evaluation, and Research, Manpower Administration, U. S. Department of Labor.

This report was prepared in the Special Projects Department in the Office of the Chief Economist under the general direction of Martin R. Gainsbrugh. The topic itself was suggested by John G. Myers, Senior Economist in the Special Projects Department, who also visualized the entire project and subsequently contributed analytical suggestions and stimulating comments. Daniel Creamer, Manager of the Special Projects Department, not only offered the encouragement which made the study feasible, but throughout provided useful criticism. Joseph Waksberg, Chief of the Statistical Methods Division of the U. S. Bureau of the Census, had pointed out several pertinent new applications which could be interestingly extended from the results propounded in the previous NICB report, and his suggestions, at different stages in the development of this project, were both timely and constructive. Alan Sunter, Coordinator of the Survey Methods Staff of the Dominion Bureau of Statistics of Canada, provided helpful advice at the time the analysis of the material was being planned. Professor Elizabeth Yen of the Department of Mathematical Statistics at Columbia University offered consistently illuminating and practical guidance in the organization of material and focusing of priorities. Comments by members of the staff of the Office of Manpower Planning, Evaluation, and Research on a preliminary draft were influential in broadening the scope of the study and in sharpening some of its conclusions.

The computations have been carried out efficiently by NICB staff members, Walter B. Brown and Barbara Feld, and the computer programs were designed by Michael Papantoniou and Judy Rosenthal under the supervision of Luke McSherry.

To all my deepest appreciation.

Although this study is neither exhaustive nor all-inclusive, it is offered with the hope that it will further the study of techniques applicable to sample design for job vacancy surveys. M.E.G.

Chapter I: INTRODUCTION AND SUMMARY

In 1967 the National Industrial Conference Board published the findings of a three-year project on the feasibility of collecting job vacancy statistics.* That study explored the problems of preparing an operational definition of a job vacancy and investigated the costs and procedures of sample survey methods, including sample designs and data collection techniques. The present study has two purposes: to analyze the reliability of vacancy estimates, and to plan efficient sample designs for job vacancy surveys.

Job vacancy statistics have important potential uses as guides to policy both nationally and in local areas. The latter category includes several different possibilities. For example, it is important to know whether or not a given local area is experiencing a shortage or a surplus of labor and to know how the situation is changing; that is, whether the local labor market is "tightening" or "loosening." Another use is to aid and identify those occupations and industries that are growing and those that are declining, so that persons concerned with job training at all levels will be better informed and thus be able to make better decisions than would be likely in the absence of such information. Also, the matching of persons and jobs would be greatly aided by an increase in information made available to both job seekers and employers in the form of numbers of vacancies by occupation and skill requirements. For all these uses, adequate, reliable statistics for each important labor market area in the nation are necessary.

Two aspects of reliability are relevant. The first is the reliability of estimates of job vacancies at a point in time (usually the survey date), both in total and for specific categories. The second is the reliability of estimated changes in job vacancies between surveys, again both in total and for selected breakdowns. A major objective of this paper is to set forth approaches to the measurement of reliability for both estimates of vacancies at a point in time and of changes between surveys. The measurement of variability was examined in The Conference Board report referred to above,* although the treatment of the variability of totals was incomplete. It did not touch at all the variability of changes. The reliability of the job vacancy rate (100 V/E+V), which is a relative measure of job vacancies, as well as changes in the rate, is also analyzed here.

---

* Measuring Job Vacancies, Studies in Business Economics No. 97, The Conference Board, 1967.

An important aspect of the feasibility of job vacancy surveys is the problem of obtaining information that is sufficiently reliable, for the uses cited above, at reasonable cost. This may be solved by preparing estimates of sample size and combining them with estimates of cost per respondent to obtain estimates of total survey costs.

The second principal objective of the present study is to explore techniques for designing efficient samples for selected areas, taking into account the prevailing job vacancy rate, the variability of job vacancies, and the costs per employer.** The variability measures and costs vary among different groups of employers. Some employer groups (with specified employment sizes and of given industries) can be identified in advance of the survey and therefore adequately enter into an efficient sample design. In this connection, the adequacy of the sample denotes that the information on numbers of vacancies and on changes in these numbers between surveys is sufficiently reliable for economic analysis.

---

* _Ibid._, Chapter 8.

** The terms "employer," "firm," "enterprise," and "establishment" are used interchangeably in this report.

Criteria for Sample Adequacy

It is difficult to choose criteria for sample adequacy in advance. Extensive experience with job vacancy statistics for policy- and decision-making in official and private use and for analysis of the value of the data as a measure of labor market activity will help to determine more precise criteria.

Some criterion must be adopted in advance, however, so that the sample size and the associated level of accuracy can be chosen. It is clearly best to state the proposed criterion explicitly, however scant the information available for choosing it. This permits discussion and evaluation of the criterion in advance of expensive data collection and may thus lead to clarification of the purposes of the survey, a refining of concepts, and in general better, more valuable data.

One procedure worthy of consideration is to determine a confidence interval for estimates of total number of job vacancies in the local area, specifying the probability of success and the width of the interval as a percentage of the estimated total. An example, to make the concept less abstract, may be taken from the original design of The Conference Board sample in Monroe County, New York (the Rochester area). The sample was designed to provide an interval estimate of the total number of job vacancies within 10% of the estimated total, with a probability of 0.95; this corresponds to a coefficient of variation of 0.05. The criterion was not satisfied in the 1965 surveys, however (see Chapter III of this report). For the purposes of studying individual industries and occupations, additional criteria of this type may be necessary, where the accuracy of subtotals of job vacancies is specified in advance.

A similar, yet statistically distinct, criterion is associated with estimation of the proportion of jobs that are vacant at a point in time. This may be referred to as the job vacancy rate and defined as the ratio of total job vacancies to the sum of total vacancies and total employment, or V/E+V. This ratio is a counterpart to the unemployment rate and is therefore more conveniently stated in percentage terms, or 100V/E+V. Reasonable criteria for accuracy of an estimate of the job vacancy rate might be constructed that are analogous to those for accuracy of an estimate of total job vacancies. Thus, we might seek to obtain, with a certain predetermined probability of success, an estimate of the job vacancy rate for an area that is within a specified per cent of the current total. For example, we might wish to obtain an estimate of the job vacancy rate that is within 10% of the actual job vacancy rate, with a probability of 0.95.

A different type of criterion is associated with the accuracy of changes in job vacancies. Here we must distinguish between changes in absolute number and changes in the job vacancy rate, for these may differ considerably according to variations in employment. Changes in the job vacancy rate indicate whether the local labor market is becoming more or less "tight," and is thus an important indicator of economic conditions in an area. Changes in the (absolute) number of job vacancies, on the other hand, are important for many decisions about training and placement.

For both types, we use criteria based on a change in the relative number of job vacancies; the criteria differ according to the base. The size of the labor market should be introduced as a base in measuring the importance of changes. Obviously a change of 500 vacancies has a different significance in a labor market such as New York City, with a labor force of more than 5.4 million (within New York State), than it has in Binghamton, with a labor force of 119,000. The criterion needed for changes in the job vacancy rate is based on the magnitude of a change that might be considered economically important. That is, the sample should be able, with a certain probability of success, to detect a change in labor market conditions that is significant in economic terms.* We believe that the detection of a change of 0.5 percentage points in the job vacancy rate in a local area would be an adequate requirement, with a probability of 0.95. A more stringent requirement would be to detect a change of 0.2 percentage points, with the same probability.

A large change in the total number of job vacancies may occur with no appreciable change in the job vacancy rate. This can happen if both total vacancies and total employment rise (or fall) in similar proportions. To establish the magnitude of absolute changes in vacancies which are considered economically significant, two possible levels of change are examined: 0.5% and 0.2% of the sum of employment and vacancies. The size of a significant change must be related to a fixed base, which must be estimated at the time the survey is being planned. This base is an estimate of the average number of jobs (employment plus vacancies) which the area surveyed will have during the survey period. The tabulation below shows the magnitude of the changes in total job vacancies, corresponding to the estimated size of the labor market area during the survey period and to specified per cent changes.

Number of Job Vacancies Corresponding to
Specified Per Cent Changes

Estimated Size of Labor Area (E+V)	Number of Vacancies Corresponding to	
	0.5 Per Cent Change	0.2 Per Cent Change
100,000	500	200
200,000	1,000	400
500,000	2,500	1,000
1,000,000	5,000	2,000

---

* This criterion was suggested by Joseph Waksberg of the U. S. Bureau of the Census.

It should be made clear that the criteria selected indicate whether an economically significant change has taken place at all, and do not serve to determine the magnitude of the change. Thus if the 0.5% criterion were adopted, and if the number of job vacancies increased by 2,000 in a labor area with 500,000 jobs, we would not be able to state, with a confidence level of 0.95, that there had been a change in the total number of vacancies; a change of at least 2,500 vacancies would be necessary to justify such a statement.

Summary of Report

The incidence of job vacancies varies widely among employers. To the extent that employers can be divided into groups, or strata, that are relatively homogeneous with respect to the incidence of job vacancies, a given degree of accuracy can be obtained with a correspondingly smaller sample (and at lower cost). The goal, therefore, is to determine those characteristics that are associated with the variability of job vacancies, to stratify the population according to those characteristics, and to select a sample separately within each stratum. If this is done successfully, the sample will be, with respect to the variability of job vacancies, relatively homogeneous within each stratum and relatively heterogeneous among strata.

Of course, practical stratification requires that the characteristics distinguishing the strata be identifiable in advance. That is, the lists from which the sample is drawn must contain information on the characteristics used for stratification. On the sampling lists, the characteristics available for employers, which are closely related to vacancies, are employment size and industry. Thus the original Conference Board sample for Rochester was stratified by employment size and industry. That sample provided data on job vacancies by employment size and by industry, for three dates in 1965: mid-February, mid-May, and mid-August, as well as on changes between these dates. A detailed description of the sample design used for the NICB surveys of Monroe County is included in Chapter II.

Computation formulas have been developed for measures of the variability of job vacancies, including numbers of job vacancies and the job vacancy rate at a point in time, and changes in these numbers and rates (Chapter III). Application of the formulas to the 1965 data provides sets of estimates of variability that should have some general applications. The degree of variability differs among the three surveys; the largest value is used here in an effort to avoid understatement of the extent of variation. The standard error of the job vacancy total in the Rochester area was 553 in the February survey, while the total proper was 7,947. This corresponds to a coefficient of variation of 0.07, and indicates that the original goal of estimating vacancies within 10% of the total value with a probability of 0.95 was not met. The indicated range for 0.95 probability is 14%. A similar result (coefficient of variation of 0.07) was found for the estimate of the job vacancy rate in February.

Also of interest are the estimates for industry groups and employment size classes. Variability, measured by the size of the standard error, was high for the following industry groups: construction; durable manufacturing; public utilities and transportation; and trade (retail and wholesale). Among employer size groups, variability was greatest for the smaller employers, those with employment of fewer than 10 workers, whether measured by size of the standard error or by the coefficient of variation.

In the Rochester sample, all employers with 250 or more employees were included, so that there is no sampling variability associated with the job vacancies of these employers. To the extent that vacancies in an industry or size group represent those of the larger employers, neither the standard error nor the coefficient of variation reflects the extent of variability. The comment about the significance of the measures cited is not intended to imply that the standard error is without value. On the contrary, the standard error indicates the precision of the results of a survey and is therefore essential to an evaluation of these results. The standard error does not, however, furnish the best guide to sample design; other measures are needed for that purpose.

The criterion suggested above for adequate sample accuracy for estimates of change is that a change of at least 0.5 percentage points in the job vacancy rate is detectable with probability of 0.95. The standard error of a change in the job vacancy rate should thus be less than 0.25 percentage points. The standard error computed from the Rochester data for the February to May change was 0.22, while for the May to August change it was 0.16. Thus the sample is of adequate size according to the 0.5 percentage point rule. The actual changes in the job vacancy rate were 0.06 points from February to May and 0.22 points from May to August, 1965.

The estimated total number of jobs (E+V) in the Rochester area was about 263,000 in February 1965 and 278,000 in May 1965. Changes of 0.5 points from February would consequently be about 1,306 vacancies, and from May about 1,390 vacancies, while changes of 0.2 points would be 526 and 556, respectively. The standard errors of the change in the total number of job vacancies were 577 for the February to May period and 449 for the May to August period. Two standard errors, corresponding approximately to 0.95 probability, are thus 1,154 and 898, respectively, for the two intervals. The Rochester sample again appears to be of adequate size for the 0.5 point rule.

The second principal objective of this study is to develop guides for sample selection at minimal cost, or optimal sample selection (Chapter IV). Here we set as one goal the selection of a sample that is of adequate precision to permit an estimate of the total number of job vacancies with a coefficient of variation of 0.025, 0.05, and 0.10. These values correspond to 95% confidence intervals for total vacancies, V', in which two standard errors are 5%, 10%, and 20% of V'.

The second goal is the selection of a sample that will detect, with a probability of 0.95, changes in job vacancies that represent a certain predetermined per cent of all jobs in the area before the change. The change we examine is 0.5 percentage points.

The relevant measure of variability of job vacancies for sample selection is the element standard error of each stratum. This is the familiar estimate of the standard deviation: in this case, the standard deviation of the job vacancies reported by the employers in a stratum.* The variability measure is combined with an appropriate cost figure to compute the optimal sample allocation among the strata. The strata are chosen on the basis of the average number and variability of vacancies, as well as the cost per employer.

Examination of the cost figures, the element standard errors, and the mean number of vacancies, computed for various industry and employment size classifications, led to the selection of 14 strata, 7 size groups cross-classified by 2 industry groups. The largest size group, 2,500 or more employees, may not be appropriate for labor market areas of substantially larger or smaller size than the one studied. In a larger area, the lower bound of the largest size group would probably be greater than 2,500, while for a smaller area, the lower bound would probably be smaller than 2,500. This can be determine by examining the distribution of employers by employment size for a specific area.

---

* This measure is multiplied by a raising factor based on the sampling fraction to obtain the standard error, used to evaluate the accuracy of an estimated total. See formula (2) in Chapter III.

The conclusions listed below are drawn from an examination of optimal sample designs computed according to the objectives stated above. They depend, to some extent, on the nature of the area studied. However, we believe that they have a large degree of applicability to areas with other industrial structures and labor market situations.

1. Stratification by size of firm and by industry improves sampling efficiency for vacancy surveys.
2. Survey costs vary significantly by size of employer in surveys seeking detailed description of each class of job vacancy. This consideration should be taken into account to obtain an efficient sample design.
3. The criteria chosen for measuring changes in job vacancies may be met with smaller and therefore less costly samples than the criteria for measuring either total vacancies or the job vacancy rate (see Chapter IV).
4. It thus appears that the detection of changes in job vacancies is more easily and less expensively achieved than the estimation of total numbers of vacancies, if the criteria stated are reasonable (see Chapter IV).

While the sample design for job vacancy surveys of a given area must take into account the characteristics of the establishments included and the prevailing job vacancy rate, it also must provide for the ultimate breakdown and the degree of reliability desired for the estimates (Chapter V). Optimal strata boundaries are functions of both the size distribution of employers and the number of employees in the largest firms. The optimal sample size for an area must be increased as the job vacancy rate diminishes. For example, in a design like the one for Rochester, where the vacancy rate was approximately 3.0, a sample size of 574 employers gives estimates of vacancies with a coefficient of variation of 0.05. However, if the vacancy rate were 1.5, a sample of 1,500 would be necessary to obtain the same level of reliability for the estimated number of vacancies. On the other hand, less stringent requirements of reliability in estimating vacancies permit reduced sample sizes. Thus, in an area like Rochester a sample of 216 employers would be sufficient to estimate vacancies with a coefficient of variation of 0.10, while one of 574 would be necessary for estimates of vacancies with a coefficient of variation of 0.05.

The methods for estimating optimal boundaries for strata and sample sizes presented in Chapter V have generalized the detailed results of Chapter IV. These methods can be applied in general to areas with characteristics differing from those found in Rochester. The specific factors which determine the optimal sample design for an area are:

1. the number of establishments (sampling units),
2. the distribution of establishments into strata according to number of employees and industry,
3. the differences among strata in vacancy variability and in survey cost per firm, and
4. the job vacancy rate.

Chapter II: STRATIFICATION AND DATA SOURCES

In this chapter, we develop a theoretical description of stratification, a feature of sample design used to classify firms and thereby improve the accuracy of statistical estimates. Although we are interested in measuring unfilled jobs, the sampling unit by necessity is the firm.* The accuracy of estimates of job vacancies is improved if firms are classified by relevant characteristics, such as employment size or type of industry, and the sample is selected within homogeneous strata.

A) Stratification in a Job Vacancy Survey

Job vacancies occur within enterprises, government, nonprofit organizations, and households. Although private households may employ people on a part-time or full-time basis, they typically have not been included in sample surveys on job vacancies for local labor areas. To include a reliable sample of households would add disproportionately to survey costs. This follows from the low incidence of households engaging employees.** Thus the sampling unit is the firm, including nonprofit organizations and government. At a point in time, firms have a given number of jobs available, some of which may be filled and some vacant. Estimates of unfilled jobs are derived from the data collected from the firms in the sample.

Certain features of sample design can reduce the cost of obtaining a desired degree of precision in the estimates. Specifically, the grouping of members of the population by characteristics closely related to (a) average size of the object of measurement, (b) variability of the object of measurement, and (c) cost per unit sampled, followed by the selection of random samples within each group, will help to produce good estimates at low cost. The classification of the population into groups, within which random samples are selected, is referred to as stratification. This method ensures the proper representation in the sample of each stratum and, therefore, reduces the variance. The variance of the stratified sample is then composed only of the variation within the strata, for the variance among strata is eliminated. To obtain the greatest gains from stratification, the analyst tries to maximize the differences among strata and to minimize the differences within strata, with respect to both variability and cost. In a survey designed to measure job vacancies, the average size and the variability of vacancies are measured by the arithmetic mean and the standard deviation of the vacancy variable, while the cost is that of the sampling unit (the firm). Since units in the population must be identified by the stratifying characteristics, factors to be used for stratification are limited to those available in the sampling frame. On occasion this constraint frustrates the attempt to choose the most relevant characteristics for stratification purposes.

---

* Although the sampling unit generally corresponds to the Census Bureau's definition of "establishment," multiple units having a central hiring office have been regarded as one sampling unit.

** For a national survey of job vacancies in households see Samuel Saben, "Regular Jobs for Household Help," Monthly Labor Review, October 1965, Vol. 88, No. 10, pp. 1,228-1,229.

The most comprehensive list of employers in a given labor area is the roster of employers covered by the state unemployment insurance law. This list has been the starting point for the NICB survey and for surveys sponsored by the Bureau of Employment Security. Thus, the NICB in its survey used the list of firms covered by the New York State Unemployment Insurance Law. This source includes a registration number, a code identifying geographic areas, size of employment each month, industry of firm, and certain data on payrolls. The geographic area code was used to select the sample of firms in Monroe County. Firms were ordered by registration number which relates to the date on which the firm joined the New York State unemployment insurance system.

Employment size and type of industry of firms are the two stratifying characteristics considered most relevant to a survey on job vacancies. Since these two classification variables are not closely related to each other, it was considered advisable to include both in the sample design.

Larger firms were expected to have a larger variance in the number of job vacancies than smaller firms. That is, a greater homogeneity is expected in the strata corresponding to smaller firms. Since for optimal stratification the sampling fraction should be proportional to the standard deviation and inversely proportional to the square root of the cost per unit in the stratum, one would use a larger sampling fraction for the strata including larger firms.* Accordingly, the sampling fraction used in the selection of the sample was varied by size of firm. The estimates of the variability of job vacancies by size of firms are given in Chapter III.

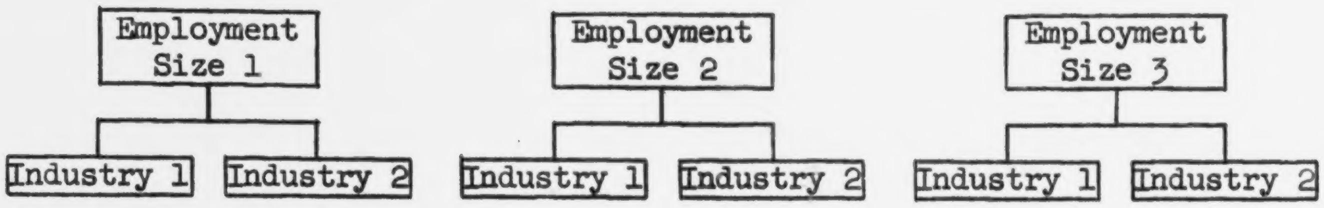
The average number and the variability of vacancies also differ markedly among industry groups. Industries that have high growth rates, or wide seasonal or cyclical variations, often have high vacancy rates. The construction industry, for example, has both high numbers of vacancies and high variability of vacancies relative to firms of the same size in other industries, during the period of seasonal upswing. Thus stratification by industry will also increase sample efficiency, as indicated by the estimates of variability given in Chapter III.

The number of strata used depends upon the accuracy of the information available on cost, average size, and variability, as well as upon the number of categories for which estimates are desired.

---

* W. G. Cochran, Sampling Techniques, Second Edition, John Wiley & Sons, Inc., 1963, pp. 95-97.

A design with stratification by size, and substratification by industry, for three sizes, each having two industry substrata, would be as follows:



A larger number of strata may increase the efficiency of the design. However, this may lead to difficulties when this procedure is carried to the extreme of including only one unit per stratum, since the variance of these strata cannot be estimated. The designs using stratification by employment size and substratification by industry are subsequently amplified and analyzed in Chapter IV, where an attempt to obtain an optimal sample design for Rochester is carried out, and in Chapter V, where a generalized method is developed.

B) Selection of the NICB Sample*

The design actually used in the 1965 surveys of Monroe County stratified firms by employment size (nine size strata were used). Within each employment size stratum, firms were ordered by industry (using 4-digit Standard Industrial Classification (S.I.C.) code), and a systematic sample was chosen from each size stratum. This design is a sample stratified by size and by industry.

The main part of the sample was selected from the list of employers paying New York State unemployment insurance payroll tax during the second quarter of 1964. However, this list did not include nonprofit institutions, government agencies, or independent professionals not covered by the New York State unemployment insurance system. Thus supplementary lists had to be developed. Another difficulty was the lack of reliable information on the variability of job vacancies for different sizes of firms. Nevertheless, an attempt was made to obtain an optimal sample. It was assumed that job vacancies followed a Poisson distribution, that the survey agency's collection cost per vacancy did not vary for the different sizes of firms, and that the job vacancy rate was constant for all firms, except for the smallest size group (i.e., firms with measure of size 1-3), where it was higher.

---

* Based on Measuring Job Vacancies.

The U.S. Census Bureau publication County Business Patterns: First Quarter 1962 included a tabulation of the number of firms by eight employment size groups in Monroe County. The following sampling fractions were determined on the basis of these data, with adjustments for the smallest size group:

Stratum	Base Employment Size	Sampling Ratio
1	0	1:100
2	1-3	1:142
3	4-7	1:52
4	8-19	1:28
5	20-49	1:12
6	50-99	1:6
7	100-249	1:3
8	250 and over	1:1

The New York State Division of Employment was requested to select from their computer records a systematic sample for each of the above size strata. The firms within each size stratum were ordered first by firm registration code. It was assumed that this ordering would approximate a random ordering of the list with respect to job vacancies in these firms. Then firms were ordered by the 4-digit S.I.C. code, within the following industry groups:

Industry Group	S.I.C. Code
1. Contract construction	15-17
2. Ordnance and durable manufacturing	19, 24, 25, 32-39
3. Nondurable manufacturing	20-23, 26-31
4. Public utilities and transportation	40-49
5. Wholesale trade	50
6. Retail trade	52-59
7. Finance, insurance, and real estate	60-67
8. Services (not including medical, legal and educational)	70-79

The Division of Employment was instructed to use a larger sampling ratio within each size group in order to provide additional firms which could be used for pretesting as well as for replacement of possible non-respondents. In addition the Division was requested to provide a complete listing of the firms in the following industries:

Industry Group	S.I.C. Code
1. Agricultural services, forestries, fisheries	01-09
2. Mining	10-14
3. Selected services (medical, legal, educational, nonprofit organizations, miscellaneous)	80-89
4. Nonclassifiable	99

After reviewing the agricultural and mining industries, which included very few firms in Monroe County, the sampling ratios corresponding to the employment size of the firms were applied to select the sample. The random number chosen did not select any firm in these categories.

For selected services (medical, legal, educational, nonprofit organizations) the first step in selecting firms in this industry group was to choose a systematic sample with the standard sampling fractions from the listing described above. Since this sampling list was incomplete, a supplementary sample of establishments not covered by the New York State Unemployment Insurance Law was chosen for the following four groups:

1. Independent professionals.

The yellow pages of the Rochester telephone directory* were scrutinized and those included in the New York State Division of Employment list were deleted. From the remaining list a systematic sample of 1:142 was selected.

2. Nonprofit organizations.

The yellow pages of the Rochester telephone directory* were reviewed and institutions listed in the "covered" employers were eliminated. The following lists were used to obtain a more complete coverage of the population: nonprofit organizations listed by the Council of Social Agencies of Rochester, a directory of hospitals in Monroe County published by the Journal of the American Hospital Association, Roman Catholic parochial schools listed by the Diocese of Rochester, and secondary schools and colleges as provided by the Rochester office of the New York State Division of Employment.

The final list was divided into two parts: employers for which some measure of employment size was available, and those for which employment size was unknown. For the former, a systematic sample corresponding to the established sampling ratios was used and for the latter, a sampling ratio of 1:50.

3. Government - Federal, state, and local.

A list of agencies was compiled from various sources. As with the nonprofit group, this list was divided into two parts: those agencies with known employment size and those with unknown employment size. The sampling ratios corresponding to the employment size of firms were used for the establishments with known employment size, and a systematic sample of 1 in 50 was used to select the sample for those with unknown employment size.

4. Public schools.

From a publication of the New York State Education Department a listing of the public schools in Monroe County was obtained. The list included a measure of employment size. The regular sampling ratios were used to select a systematic sample.

---

* The Rochester metropolitan telephone directory covers all of Monroe County and some areas outside this county. If a unit selected was outside Monroe County, it was excluded from the sample.

After the February 1965 survey was carried out, a supplementary sample was added to the original one*. Eleven firms were reported "out of business" in February. They were identified under their respective major industry groups and a systematic sample of 11 new firms which had started operations after the second quarter of 1964 was selected. These firms were selected to correspond to the industry of the firms "out of business," but since no employment information was available at that time, a sampling ratio of 1:50 corresponding to the stratum of employers with unknown employment size was used.

The sample included approximately 400 employers in Monroe County. These employers reported (in February, May, and August, 1965) the number of employees in their respective firms and described the jobs they were seeking to fill, specifying their requirements as to occupation, experience, education, and sex, as well as the desirable starting date. Estimates of job vacancies, of job vacancy rates, and of changes in these quantities were derived from these reports.

Summary

In the light of present information the sample design of a job vacancy survey should include a stratification of the firms by employment size and type of industry. The specification of the strata to be used for designing the sample should be based upon data showing how job vacancies vary according to the characteristics of the firms and should take into account the size of the labor force area and the distribution of firms by strata. The strata should group firms homogeneous with regard to job vacancies and maximize the differences between firms in separate strata.

The NICB surveys of Monroe County used nine employment size strata and within each size stratum subclassified firms by type of industry.

---

* The vacancy rate in these supplementary firms was 12.5, as compared with 3.2 for all employers (Measuring Job Vacancies, p. 143). Therefore, the updating of samples is highly desirable for a vacancy survey. However, delays in the registration of new firms sometimes make it difficult to maintain an up-to-date sampling frame.

The strata used in this survey may or may not have produced the most accurate results for this area. Even if the strata used for the sample design were optimal for Monroe County, in order to apply the procedure to other areas, the specific characteristics of the area must be taken into consideration. Since the principal purpose is to make possible the designing of sample surveys to collect job vacancy statistics for differing labor force areas, the general features of sample design must be evaluated. The different circumstances of each particular area for which a sample design is to be made will determine the stratification necessary.

In order to be able to determine an adequate stratification, we shall go on to analyze the variability of the sample estimates of Monroe County as an example of the method employed.

Chapter III: MEASURES OF VARIABILITY

When the variability of sample estimates by employment size and by type of industry of firms is to be analyzed, it is advisable to discuss the computation of the following estimates:

a) total number of job vacancies,

b) job vacancy rates,

c) changes in total number of job vacancies between survey periods, and

d) changes in job vacancy rates between survey periods.

For each of these estimates, the standard error is used to evaluate the variability of the sample used in the NICB survey of Monroe County.

However, a simpler sample design will be used to estimate the variance of the sample estimates. That is, the variance is computed for a sample stratified by size of firm only, on the supposition that within each size stratum a simple random sample of firms was selected. This procedure simplifies computations, but leaves out of consideration the fact that firms were classified by industry within each size stratum and that from each stratum a systematic sample was selected. The approximation of the variance gives an estimate somewhat larger than the true variance of the design. Therefore, if the calculated variance is used to compute the reliability of the sample estimates, the results will be on the conservative side.

Although nonsampling errors (both those increasing the sampling variances and those creating biases) can significantly affect results, the sample design used in the NICB vacancy surveys does not allow for such estimation.* The measures of variability available from the NICB surveys permit estimates of sampling variability only. To estimate nonsampling errors a replicated sample must be used. However, the advantages of replicated sample designs are sometimes offset by other factors. If only two replications are carried out, estimates of variance are quite poor; on the other hand, for multiple replications, each replicate must comprise a large number of units; this enlarges the total sample size and significantly increases costs. The job vacancy surveys being carried out by the Dominion Bureau of Statistics in Canada use a replicated sample design. However, this is still an incipient project to the extent that reports giving estimates of total variability of vacancy estimates have not yet been published.

---

* A discussion of nonsampling errors in the NICB surveys is included in Measuring Job Vacancies, Chapter 7.

It should also be mentioned that if a large percentage of sample firms do not respond, estimates based on the respondent firms may be biased, since the characteristics of firms that do respond are not necessarily the same as those of nonrespondent firms. In the 1965 NICB surveys the problem of nonresponse was negligible, since 99% of the sample firms responded.

The tables that follow specify (a) the estimates, (b) the standard error of these estimates, and (c) the coefficient of variation. The latter, the ratio of the standard error to the quantity estimated, gives a measure of the reliability of the estimates. For example, the coefficient of variation of the total number of job vacancies in February 1965 is 0.07, corresponding to a standard error of 553 job vacancies. This indicates that if many random samples of employers had been used in the February 1965 survey, the 95% confidence interval for total job vacancies would include the true population value in 95% of the samples. The 95% confidence interval for total job vacancies in the NICB February 1965 survey is 7,947 plus or minus 1,106, i.e., between 6,841 and 9,053 job vacancies. In this example the analyst would infer that the sample was large enough to yield reliable estimates of total number of job vacancies. The same logic applies to the reliability of the estimated change in job vacancies between survey dates or between job vacancy rates. The latter is defined as the ratio of job vacancies to the total of employed plus job vacancies, i.e., the total demand for labor, satisfied and unsatisfied.

Before designing a sample, an analyst must decide what magnitude of changes in job vacancies has economic significance for his particular purpose. For example, it might be significant to detect changes in vacancies which represent 0.5% of total labor demand (employment plus vacancies) in Monroe County. Half of one per cent of 276,000 is 1,380 job vacancies, and a standard error of 690 would be sufficient to obtain 95% confidence in detecting this change. The estimate of change between February and May 1965 may serve as an illustration - it is 829 vacancies, and the estimated standard error is 557; in this case, the coefficient of variation is 0.7. This means that the estimate of change is not reliable in terms of the standard error, but, as noted at the outset, such a relatively small change may be considered of little economic importance.

A) Actual Measures of Variability for Each Survey Period

1) Vacancies by industry

The employment size strata and sampling fractions used for the selection of the sample are listed below. For most employers the employment size corresponds to that in the records of the New York State unemployment insurance system for June 30, 1964, hereafter referred to as the selection date.

The employment size strata on the selection date are listed below. Hereafter, the letter j will refer to these strata (j = 1, 2, 3, ..., 9).

Stratum (j)	Base Employment Size	Sampling Ratio
1	0	1:100
2	1-3	1:142
3	4-7	1:52
4	8-19	1:28
5	20-49	1:12
6	50-99	1:6
7	100-249	1:3
8	250 and over	1:1
9	unknown	1:50

Table 1 presents the estimated total number of job vacancies, classified by industry for the three surveys (February, May, and August, 1965) and subsequently will be referred to as Survey 1, 2, or 3. However, in formulas the survey number will be identified by the letter r (r = 1, 2, 3). The industry groups defined in terms of the Standard Industrial Classification (S.I.C.) of the firms are listed below. The letter k will be used to identify an industry group (k = 1, 2, 3, ..., 9).

Industry Group (k)	S.I.C.
1. Durable manufacturing	19, 24-25, 32-39
2. Nondurable manufacturing	20-23, 26-31
3. Construction	15-17
4. Public utilities	40-49
5. Trade, retail and wholesale	50-59
6. Finance, insurance, and real estate	60-69
7. Services, excluding education	70-81, 83-89
8. Education	82
9. Government	91-94

The letter i will refer to an individual firm.

Let

$$V^{kr}_{ji} = \begin{cases} V^{r}_{ji} & \text{Vacancies on survey } r \text{ of firm } i \text{ in size stratum } j \text{ on the selection date, and in industry } k. \\ 0 & \text{For firms not in industry } k. \end{cases}$$

N_j^o : Number of firms in the population of size stratum *j* on the selection date.

n_j^o : Number of firms in the sample of size stratum *j* on the selection date.

$f_j = \frac{n_j^o}{N_j^o}$ Sampling fraction in the original size stratum *j*.

At the time of the first survey some firms selected for inclusion in the sample were out of business. Before the second survey a supplementary sample was selected. Employment size information was not available for these new firms, so they were placed in size stratum 9. These firms will be denoted as n_9' . Certain firms which responded to the first survey, were out of business or refused to answer on the second or third survey. These firms will be denoted as n_j^r , where *r* = 2, 3. Therefore, the total number of firms in the sample in stratum *j* on the first survey is n_j^o, and on the second or third survey they are $n_j^o - n_j^r$, if $j \neq 9$, and $n_9^o + n_9' - n_9^r$ for *j* = 9. The corresponding estimates for the population are N_j^o , for the first survey, and $N_j^o - N_j^r$, if *j* ≠ 9, and $N_9^o + N_9' - N_9^r$ for *j* = 9 for the second and third surveys.

Throughout Part A, formulas are written based on the first survey. The estimate of total vacancies for an industry *k* is:

$$V^{k1} = \sum_{j=1}^{9} \frac{1}{f_j} \sum_{i=1}^{n_j^o} V_{ji}^{k1} \qquad (1)$$

The standard error of the vacancies is the square root of the variance. The variance of the vacancies in an industry group k is:

$$\mathrm{Var}\,(V^{k1}) = \sum_{j=1}^{q}\left[\frac{(N_j^o)^2}{n_j^o}\left(1 - f_j\right)\right]\frac{\sum_{i=1}^{n_j^o}\left(V_{ji}^{k1}\right)^2 - \frac{\left(\sum_{i=1}^{n_j^o} V_{ji}^{k1}\right)^2}{n_j^o}}{n_j^o - 1} \qquad (2)$$

To adapt the above formula for Survey 2 or Survey 3, n_j^o should be modified by adding the supplementary sample (n'_q), and by subtracting the firms which did not respond on these surveys (n_j^r); similarly N_j^o should be modified; and the vacancies would refer to Survey 2 or Survey 3.

To estimate the total number of vacancies the following computation should be made:

$$V^1 = \sum_{j=1}^{q} \frac{1}{f_j} \sum_{i=1}^{n_j^o} V_{ji}^1 \qquad , \qquad (3)$$

where

V_{ji}^r: Number of vacancies on Survey r, of firm i in size stratum j on the selection date.

The estimate of the variance of the total is:

$$\mathrm{Var}\,(V^1) = \sum_{j=1}^{q}\left[\frac{(N_j^o)^2}{n_j^o}\left(1 - f_j\right)\right]\frac{\sum_{i=1}^{n_j^o}\left(V_{ji}^1\right)^2 - \frac{\left(\sum_{i=1}^{h_j^o} V_{ji}^1\right)^2}{h_j^o}}{h_j^o - 1} \quad . \qquad (4)$$

Table 1. Total Number of Job Vacancies, Standard Error, and Coefficient of Variation, by Industry Group

Survey Period and Industry Group	Job Vacancies (1)	Standard Error of Vacancies (2)	Coefficient of Variation of Vacancies (2)÷(1) (3)
February			
Durable manufacturing	2,590	195.0	0.075
Nondurable manufacturing	431	27.9	0.065
Construction	850	347.0	0.408
Public utilities and transportation	408	284.0	0.696
Trade, retail and wholesale	1,004	233.0	0.232
Finance, insurance, and real estate	294	106.0	0.361
Services, excluding education	710	162.0	0.228
Education, public and private	1,320	82.4	0.062
Government	340	43.9	0.129
Total	7,947	553.0	0.070
May			
Durable manufacturing	3,418	289.0	0.085
Nondurable manufacturing	804	140.0	0.174
Construction	836	270.0	0.323
Public utilities and transportation	257	101.0	0.393
Trade, retail and wholesale	1,296	241.0	0.186
Finance, insurance, and real estate	224	92.2	0.412
Services, excluding education	856	147.0	0.172
Education, public and private	723	64.4	0.089
Government	362	67.1	0.185
Total	8,776	469.0	0.053
August			
Durable manufacturing	3,516	312.0	0.089
Nondurable manufacturing	1,057	149.0	0.141
Construction	816	334.0	0.409
Public utilities and transportation	181	83.6	0.462
Trade, retail and wholesale	1,284	242.0	0.188
Finance, insurance, and realestate	179	52.7	0.294
Services, excluding education	792	209.0	0.264
Education, public and private	386	35.5	0.092
Government	357	59.6	0.167
Total	8,568	553.0	0.065

Source: Based on data from all units that responded to the NICB surveys.

The results of applying these formulas to the Monroe County survey data are summarized in Table 1. It is found that the construction industry has the highest standard error in two of the three surveys. Its coefficient of variation is above 0.3 in all three surveys. The high variability of this industry is probably a reflection of its seasonal nature owing to climate. Durable manufacturing and trade are industries with a large standard error also. This finding is relevant to any attempt to design an optimal sample. To obtain the greatest gains possible, firms in construction, trade, and durable industries are grouped in Chapter IV (dealing with optimal sample design) to form a special stratum having a higher standard error.

2) Vacancies by size

To measure the variability of the estimates by employment size of respondents, the procedure is as indicated below. The letter m is used to identify the present employment size of a firm ($m = 1, 2, 3, \ldots, 8$). Firms are grouped into the following employment sizes:

Firms Grouped by Employment Size (m)	Number of Employees
1	0-9
2	10-19
3	20-49
4	50-99
5	100-249
6	250-999
7	1,000-2,499
8	2,500 and over

The letters (i, j, r) are used as defined previously.

The definition of the vacancy variable to be used for estimating the characteristics corresponding to employment size m is:

$$V_{ji}^{mr} = \begin{cases} V_{ji}^{r} & \text{Vacancies of firm } i \text{ in stratum } j \text{ on the selection date, and in employment size } m \text{ at survey } r. \\ 0 & \text{Firms not in group } m \text{ at survey } r. \end{cases}$$

The estimate of total number of vacancies in size group m on Survey 1 is:

$$V^{m1} = \sum_{j=1}^{4} \frac{1}{f_j} \sum_{i=1}^{n_j^\circ} V_{ji}^{m1} \tag{5}$$

The variance of this estimate is:

$$\mathrm{Var}\,(V^{m1}) = \sum_{j=1}^{4} \left[\frac{(N_j^\circ)^2}{n_j^\circ} (1 - f_j) \right] \frac{\sum_{i=1}^{n_j^\circ} \left(V_{ji}^{m1} \right)^2 - \frac{\left(\sum_{i=1}^{n_j^\circ} V_{ji}^{m1} \right)^2}{n_j^\circ}}{n_j^\circ - 1} \tag{6}$$

To modify the above formula for Surveys 2 or 3, the vacancies (V_{ji}^{mr}) correspond to $r = 2, 3$; n_j° and N_j° are modified to include the firms added to the sample before Survey 2, and to exclude the firms which did not respond on Surveys 2 or 3.

The standard error is largest for the smallest size group of 0 - 9 employees (Table 2). The largest size groups of 1,000 - 2,499 employees, and of 2,500 and over have zero standard error, since all of these firms were included in the sample. Firms having 250 to 999 employees on the selection date were also included with certainty; however, the standard error for this size class is not zero, because

Table 2. Total Number of Job Vacancies, Standard Error, and Coefficient of Variation, by Employment Size

Survey Period and Employment Size	Job Vacancies (1)	Standard Error of Vacancies (2)	Coefficient of Variation of Vacancies (2)÷(1) (3)
February			
0-9	1,126	415.0	0.369
10-19	420	122.0	0.290
20-49	722	185.0	0.256
50-99	672	290.0	0.432
100-249	627	157.0	0.250
250-999	1,273	17.1	0.013
1,000-2,499	591	0	0
2,500 and over	2,516	0	0
Total	7,947	553.0	0.070
May			
0-9	784	246.0	0.314
10-19	786	198.0	0.252
20-49	1,172	288.0	0.246
50-99	834	195.0	0.234
100-249	701	184.0	0.262
250-999	1,420	79.2	0.056
1,000-2,499	643	0	0
2,500 and over	2,436	0	0
Total	8,776	469.0	0.053
August			
0-9	890	349.0	0.392
10-19	848	322.0	0.380
20-49	884	246.0	0.278
50-99	707	137.0	0.194
100-249	853	167.0	0.196
250-999	1,016	37.1	0.037
1,000-2,499	824	0	0
2,500 and over	2,546	0	0
Total	8,568	553.0	0.065

Source: Based on data from all units that responded to the NICB surveys.

some firms which had less than 250 employees when the sample was selected, had more than 250 employees when the survey data were collected.

3) Vacancy rate by industry

The job vacancy rate is ($\frac{100V}{V+E}$) (V: job vacancies; E: employment). Let k identify the industry group (k = 1, 2, 3, ..., 9). The variable V_{ji}^{kr} has already been defined. Total labor demand may now be defined as:

$$(V_{ji} + E_{ji})^{kr} = \begin{cases} (V_{ji} + E_{ji})^{r} & \text{Vacancies plus employment on survey } r \text{ of firm } i \text{ in stratum } j \text{ on the selection date, and in industry } k. \\ 0 & \text{For firms not in industry } k. \end{cases}$$

The vacancy rate for industry k on Survey 1 is:

$$100\left(\frac{V}{V+E}\right)^{k1} = \frac{100\,V^{k1}}{(V+E)^{k1}} = \frac{100\sum_{j=1}^{9}\frac{1}{f_j}\sum_{i=1}^{n_j^0} V_{ji}^{k1}}{\sum_{j=1}^{9}\frac{1}{f_j}\sum_{i=1}^{n_j^0}(V_{ji}+E_{ji})^{k1}} \tag{7}$$

To compute the variance of the vacancy rate, a first approximation formula is used.* The variance of the vacancy rate of industry k on survey r is:

$$\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{k1}\right] = \left[\frac{100\,V^{k1}}{(V+E)^{k1}}\right]^2 \left\{ \frac{\mathrm{Var}\,(V^{k1})}{(V^{k1})^2} + \frac{\mathrm{Var}\left[(V+E)^{k1}\right]}{\left[(V+E)^{k1}\right]^2} - \frac{2\,\mathrm{Cov}\left[V^{k1}, (V+E)^{k1}\right]}{V^{k1}\left[(V+E)^{k1}\right]} \right\}, \tag{8}$$

---

*Leslie Kish, Survey Sampling, John Wiley & Sons, Inc., New York, 1965, p. 207.

where $\mathrm{Var}\,(V^{k1})$ is computed by formula (2),

$$\mathrm{Var}\left[(V+E)^{k1}\right] = \sum_{j=1}^{9}\left[\frac{(N_j^o)^2}{n_j^o}(1-f_j)\right]\frac{\sum_{i=1}^{n_j^o}\left[(V_{ji}+E_{ji})^{k1}\right]^2 - \frac{\left[\sum_{i=1}^{n_j^o}(V_{ji}+E_{ji})^{k1}\right]^2}{n_j^o}}{n_j^o - 1}, \quad (9)$$

and

$$\mathrm{Cov}\left[V^{k1}, (V+E)^{k1}\right] = \sum_{j=1}^{9}\left[\frac{(N_j^o)^2}{n_j^o}(1-f_j)\right]\frac{\sum_{i=1}^{n_j^o} V_{ji}^{k1}\left[(V_{ji}+E_{ji})^{k1}\right] - \frac{\sum_{i=1}^{n_j^o} V_{ji}^{k1}\left[\sum_{i=1}^{n_j^o}(V_{ji}+E_{ji})^{k1}\right]}{n_j^o}}{n_j^o - 1}. \quad (10)$$

To estimate the total vacancy rate on Survey 1, we use:

$$100\left(\frac{V}{V+E}\right)^1 = \frac{100\,V^1}{(V+E)^1} = \frac{100\sum_{j=1}^{9}\frac{1}{f_j}\sum_{i=1}^{n_j^o} V_{ji}^1}{\sum_{j=1}^{9}\frac{1}{f_j}\sum_{i=1}^{n_j^o}(V_{ji}+E_{ji})^1} \quad (11)$$

The variance of the total vacancy rate on Survey 1 is estimated by:

$$\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^1\right] = \left[100\left(\frac{V}{V+E}\right)^1\right]^2\left\{\frac{\mathrm{Var}(V)^1}{(V^1)^2} + \frac{\mathrm{Var}\,(V+E)^1}{[(V+E)^1]^2} - \frac{2\,\mathrm{Cov}[V^1, (V+E)^1]}{V^1[(V+E)^1]}\right\}, \quad (12)$$

Table 3. Job Vacancy Rate, Standard Error, and Coefficient of Variation, by Industry Group

Survey Period and Industry Group	Job Vacancy Rate $\frac{100V}{V+E}$ (1)	Standard Error of Vacancy Rate (2)	Coefficient of Variation of Vacancy Rate (2)÷(1) (3)
February			
Durable manufacturing	2.62	0.172	0.066
Nondurable manufacturing	1.73	0.124	0.072
Construction	6.94	2.325	0.335
Public utilities and transportation	4.09	2.458	0.601
Trade, retail and wholesale	2.31	0.485	0.210
Finance, insurance, and real estate	2.91	0.804	0.276
Services, excluding education	2.46	0.547	0.222
Education, public and private	6.09	0.182	0.030
Government	2.60	0.282	0.108
Total	3.02	0.198	0.066
May			
Durable manufacturing	3.37	0.249	0.074
Nondurable manufacturing	2.99	0.468	0.157
Construction	4.94	1.426	0.289
Public utilities and transportation	2.58	0.852	0.330
Trade, retail and wholesale	2.85	0.458	0.161
Finance, insurance, and real estate	2.31	0.831	0.360
Services, excluding education	2.56	0.472	0.184
Education, public and private	3.46	0.176	0.051
Government	2.76	0.379	0.137
Total	3.16	0.167	0.053
August			
Durable manufacturing	3.35	0.258	0.077
Nondurable manufacturing	3.75	0.492	0.131
Construction	4.72	1.822	0.386
Public utilities and transportation	1.80	0.752	0.418
Trade, retail and wholesale	2.73	0.440	0.161
Finance, insurance, and real estate	1.66	0.446	0.269
Services, excluding education	2.39	0.625	0.262
Education, public and private	1.82	0.103	0.057
Government	2.49	0.359	0.144
Total	2.98	0.187	0.063

Source: Based on data from all units that responded to the NICB surveys.

where Var V^1 is computed using formula (4),

$$\mathrm{Var}\ (V+E)^1 = \sum_{j=1}^{9}\left[\frac{(N_j^o)^2}{n_j^o}(1-f_j)\right] \frac{\sum_{i=1}^{n_j^o}\left[(V_{ji}+E_{ji})^1\right]^2 - \frac{\left[\sum_{i=1}^{n_j^o}(V_{ji}+E_{ji})^1\right]^2}{n_j^o}}{n_j^o - 1} , \qquad (13)$$

and

$$\mathrm{Cov}\left[V^1, (V+E)^1\right] = \sum_{j=1}^{9}\left[\frac{(N_j^o)^2}{n_j^o}(1-f_j)\right] \frac{\sum_{i=1}^{n_j^o} V_{ji}^1\left[(V_{ji}+E_{ji})^1\right] - \frac{\left(\sum_{i=1}^{n_j^o} V_{ji}^1\right)\left[\sum_{i=1}^{n_j^o}(V_{ji}+E_{ji})^1\right]}{n_j^o}}{n_j^o - 1} . \qquad (14)$$

The standard error of the vacancy rate (Table 3) is largest for the construction industry in two of the three surveys, as it was for the estimate of total job vacancies. The coefficient of variation is higher for firms in the construction, public utilities, and trade industries than for firms in other industries. The coefficient of variation for the total job vacancy rate varies between 0.05 and 0.07.

4) Vacancy rate by size

Table 4 gives the job vacancy rate and the estimated standard error by employment size of firms. The estimation procedure is similar to the one described in Table 3, except that firms are classified by the employment size groups used in Table 2. That is, instead of classifying by industry groups ($\underline{k}$ = 1, 2, 3, ... 9), employment sizes, designated by variable $\underline{m}$ ($\underline{m}$ = 1, 2, 3, ..., 8), are used to classify firms.

The coefficient of variation is largest for the smallest firms (0-9 employees), except for Survey 1 in which the firms with 50-99 employees had a larger coefficient of variation. Firms with 1,000 employees or more again had zero standard errors, because all firms in the population were included in the sample.

Table 4. Job Vacancy Rate, Standard Error, and Coefficient of Variation, by Employment Size

Survey Period and Employment Size	Job Vacancy Rate $\frac{100V}{V+E}$ (1)	Standard Error of Vacancy Rate (2)	Coefficient of Variation of Vacancy Rate (2)÷(1) (3)
February			
0-9	3.66	1.266	0.346
10-19	2.81	0.699	0.249
20-49	3.56	0.806	0.226
50-99	4.13	1.524	0.369
100-249	2.79	0.544	0.195
250-999	2.84	0.045	0.016
1,000-2,499	2.00	0	0
2,500 and over	2.99	0	0
Total	3.02	0.198	0.066
May			
0-9	2.43	0.743	0.306
10-19	4.47	0.993	0.222
20-49	5.64	1.032	0.183
50-99	3.89	1.104	0.284
100-249	2.53	0.580	0.229
250-999	3.01	0.101	0.034
1,000-2,499	2.51	0	0
2,500 and over	2.86	0	0
Total	3.16	0.167	0.053
August			
0-9	2.72	1.028	0.378
10-19	4.80	1.712	0.357
20-49	4.00	0.911	0.228
50-99	3.38	0.777	0.230
100-249	3.03	0.493	0.163
250-999	2.20	0.051	0.023
1,000-2,499	3.01	0	0
2,500 and over	2.76	0	0
Total	2.98	0.187	0.063

Source: Based on data from all units that responded to the NICB surveys.

B) Measures of Variability for Changes Between Survey Periods

Tables 5, 6, 7, and 8 are based on the 393 sample firms which responded to all three surveys. The firms for which data were available for only one or two surveys were excluded.

1) Changes in vacancies by industry

In Table 5 firms are classified into the same industry groups used in Tables 1 and 3, and size groups have been collapsed. Subscript k is used to identify these industry groups (k = 1, 2, 3, ...,9). In estimating changes, two survey periods will be compared. These periods are labeled as r_1 and r_2, where $r_1 \neq r_2$ and $r_1 = 1,2$, $r_2 = 2,3$.

The absolute value of the difference in vacancies is:

$$|V^{kr_1} - V^{kr_2}| ,$$

where V^{kr_1} and V^{kr_2} are estimated by formula (1).

The variance of the difference is estimated as follows:

$$\mathrm{Var}\,(V^{kr_1} - V^{kr_2}) = \mathrm{Var}\,(V^{kr_1}) + \mathrm{Var}\,(V^{kr_2}) - 2\,\mathrm{Cov}\,(V^{kr_1}, V^{kr_2}) , \quad (15)$$

where Var (V^{kr_1}) and Var (V^{kr_2}) are estimated with formula (2). A firm in industry k in survey r_1, will still be in industry k in survey r_2. To estimate the covariance term it may be recalled that:

$$V_{ji}^{kr_1} = \begin{cases} V_{ji}^{r_1} & \text{Vacancies in survey } r_1 \text{ of firms in industry } k. \\ 0 & \text{For firms not in industry } k. \end{cases}$$

Table 5. **Changes in Job Vacancies, Standard Error,** and Coefficient of Variation, by Industry Group

Survey Period and Industry Group	Absolute Value of Change in Job Vacancies (1)	Standard Error of Change in Vacancies (2)	Coefficient of Variation of Change in Vacancies **(2)÷(1)** **(3)**
February - May			
Durable manufacturing	728	164	0.23
Nondurable manufacturing	249	81	0.33
Construction	114	324	2.84
Public utilities and transportation	151	299	1.98
Trade, retail and wholesale	292	247	0.85
Finance, insurance, and real estate	70	48	0.69
Services, excluding education	146	199	1.36
Education	597	59	0.10
Government	22	56	2.55
Total	505	577	1.14
May - August			
Durable manufacturing	148	211	1.43
Nondurable manufacturing	377	85	0.23
Construction	270	241	0.89
Public utilities and transportation	76	58	0.76
Trade, retail and wholesale	12	206	17.17
Finance, insurance, and real estate	45	66	1.47
Services, excluding education	64	201	3.14
Education	337	30	0.09
Government	5	37	7.40
Total	284	449	1.58
February - August			
Durable manufacturing	876	237	0.27
Nondurable manufacturing	626	141	0.23
Construction	384	366	0.95
Public Utilities and transportation	227	296	1.30
Trade, retail and wholesale	280	219	0.78
Finance, insurance, and real estate	115	83	0.72
Services, excluding education	82	261	3.18
Education	934	66	0.07
Government	17	72	4.24
Total	221	658	2.98

Source: Based on the 393 units that responded to all three NICB surveys.

$$V_{ji}^{kr_2} = \begin{cases} V_{ji}^{r_2} & \text{Vacancies in survey } \underline{r_2} \text{ of firms in industry } \underline{k}. \\ 0 & \text{For firms not in industry } \underline{k}. \end{cases}$$

Since we have limited the analysis to firms which responded to all three surveys:

$N_j^{r_1} = N_j^{r_2} = N_j^{\circ} - N_j' = N_j$, where N_j° are the firms in size stratum j of the population on the selection date, and N_j' corresponds to the firms for which data were not available on all three surveys. N_j' is estimated as n_j'/f_j. The population is reduced by the number of firms which did not respond to all surveys.

The firms in the sample are: $n_j^{r_1} = n_j^{r_2} = n_j^{\circ} - n_j' = n_j$. $n_j \leq n_j^{\circ}$ since the firms which did not respond to all surveys (n_j') are omitted. Then the covariance term is:

$$\text{Cov}\ (V^{kr_1}, V^{kr_2}) = \sum_{j=1}^{q}\left[\frac{N_j^2}{n_j}(1-f_j)\right] \frac{\sum_{i=1}^{n_j} V_{ji}^{kr_1} V_{ji}^{kr_2} - \frac{\left(\sum_{i=1}^{n_j} V_{ji}^{kr_1}\right)\left(\sum_{i=1}^{n_j} V_{ji}^{kr_2}\right)}{n_j}}{n_j - 1} \quad . \quad (16)$$

The absolute value of the difference of the total number of vacancies is estimated as below:

$$|V^{r_1} - V^{r_2}| \quad ,$$

where V^{r_1} and V^{r_2} are computed using formula (3).

The variance of the difference in the total number of vacancies is computed by:

$$\text{Var}(V^{r_1} - V^{r_2}) = \text{Var}\ (V^{r_1}) + \text{Var}\ (V^{r_2}) - 2\ \text{Cov}\ (V^{r_1}, V^{r_2}) \quad , \quad (17)$$

where Var (V^{r_1}) and Var (V^{r_2}) are estimated by formula (4). The covariance is calculated by:

$$\text{Cov}\ (V^{r_1}, V^{r_2}) = \sum_{j=1}^{q}\left[\frac{N_j^2}{n_j}(1-f_j)\right] \frac{\sum_{i=1}^{n_j} V_{ji}^{r_1} V_{ji}^{r_2} - \frac{\left(\sum_{i=1}^{n_j} V_{ji}^{r_1}\right)\left(\sum_{i=1}^{n_j} V_{ji}^{r_2}\right)}{n_j}}{n_j - 1} \quad , \quad (18)$$

where

$V_{ji}^{r_1}$: Number of vacancies in survey r_1 of firm i in size stratum j on the selection date.

$V_{ji}^{r_2}$: Number of vacancies in survey r_2 of firm i in size stratum j on the selection date.

The standard error of the changes in total job vacancies is larger than the estimated size of this change and this was the case for about half of the industry groups. Therefore, the NICB sample of Monroe County is not big enough to pinpoint the small changes in job vacancies which occurred between the three 1965 surveys, but these changes may be considered economically insignificant. Table 5 includes estimates of changes that occurred in three-month periods and in a six-month period.

2) Changes in vacancies by size

Table 6 gives estimates changes of vacancies by employment size of firms. The 393 firms for which data for the three surveys are available are grouped to form the employment size classification used, denoted by the letter s (s = 1, 2, ..., 5):

Employment Size of Firm (s)	Number of Employees
1	0-9 employees in all three surveys
2	10-49 employees in all three surveys
3	50-249 employees in all three surveys
4	250 or more employees in all three surveys
5	Firms which do not correspond to groups 1-4 as defined above. This group includes 39 firms which shift employment size groups between survey periods

Table 6. **Changes in Job Vacancies, Standard Error,** and Coefficient of Variation, by Employment Size

Survey Period and Employment Size	Absolute Value of Change in Job Vacancies (1)	Standard Error of Change in Vacancies (2)	Coefficient of Variation of Change in Vacancies (2)÷(1) (3)
February - May			
Firms that remain in same employment size			
0-9	298	387.00	1.30
10-49	338	198.00	0.59
50-249	116	336.00	2.90
250 and over	175	58.80	0.34
Firms that change employment size	174	171.00	0.98
Total	505	577.00	1.14
May-August			
Firms that remain in same employment size			
0-9	66	215.00	3.26
10-49	112	210.00	1.88
50-249	99	226.00	2.28
250 and over	121	61.20	0.51
Firms that change employment size	110	237.00	2.15
Total	284	449.00	1.58
February - August			
Firms that remain in same employment size			
0-9	364	375.00	1.03
10-49	450	320.00	0.71
50-249	17	324.00	19.06
250 and over	54	2.45	0.05
Firms that change employment size	64	288.00	4.50
Total	221	658.00	2.98

Source: Based on the 393 units that responded to all three NICB surveys.

The vacancies of employment size s are defined as follows:

$$V_{ji}^{sr_1} = \begin{cases} V_{ji}^{r_1} & \text{Vacancies on survey } r_1 \text{ of firm } i \text{ in size stratum } j \text{ on the selection date, and in size group } s \text{ on survey } r_1. \\ 0 & \text{For firms not in size group } s \text{ on survey } r_1. \end{cases}$$

$$V_{ji}^{sr_2} = \begin{cases} V_{ji}^{r_2} & \text{Vacancies on survey } r_2 \text{ of firm } i \text{ in size stratum } j \text{ on the selection date, and in size group } s \text{ on survey } r_2. \\ 0 & \text{For firms not in size group } s \text{ on survey } r_2. \end{cases}$$

If a firm is in a particular size group s on survey r_1, it will be in the same size group on survey r_2.

The absolute value of the difference of vacancies between two survey periods is:

$$| V^{sr_1} - V^{sr_2} | \quad , \tag{19}$$

where

$$V^{sr_1} = \sum_{j=1}^{q} \frac{1}{f_j} \sum_{i=1}^{n_j} V_{ji}^{sr_1} \quad . \tag{20}$$

The variance of the difference is:

$$\text{Var } (V^{sr_1} - V^{sr_2}) = \text{Var } V^{sr_1} + \text{Var } V^{sr_2} - 2 \text{ Cov } (V^{sr_1}, V^{sr_2}) \quad , \tag{21}$$

$$\text{where Var } V^{sr_1} = \sum_{j=1}^{q} \left[\frac{N_j^2}{n_j}(1-f_j)\right] \frac{\sum_{i=1}^{n_j} (V_{ji}^{sr_1})^2 - \frac{\left(\sum_{i=1}^{n_j} V_{ji}^{sr_1}\right)^2}{n_j}}{n_j - 1} \quad . \tag{22}$$

$\text{Var}(V^{sr_2})$ is computed with an equation similar to formula (22), but using variable $V_{ji}^{sr_2}$. The covariance term is estimated as follows:

$$\text{Cov } (V^{sr_1}, V^{sr_2}) = \sum_{j=1}^{q} \left[\frac{N_j^2}{n_j}(1-f_j)\right] \frac{\sum_{i=1}^{n_j} V_{ji}^{sr_1} V_{ji}^{sr_2} - \frac{\left(\sum_{i=1}^{n_j} V_{ji}^{sr_1}\right)\left(\sum_{i=1}^{n_j} V_{ji}^{sr_2}\right)}{n_j}}{n_j - 1} \quad . \tag{23}$$

Of the 39 firms in group 5, 21 firms shift between size group 1 and 2; seven firms shift between size group 2 and 3; seven shift between size group 3 and 4; three shift between size group 1 and 3, and one firm shifts from group 1, to 2, and then to 3. Some of these firms may have seasonal fluctuations in their employment; others may be expanding or shrinking their employment.

Although the following results may have little economic importance, the standard error of the difference in vacancies given in Table 6 is of the same order of magnitude as the absolute value of the difference in vacancies for all size groups, except for the firms with 250 employees or more; therefore, the NICB sample survey is not geared to detect accurately small changes in vacancies which actually occurred.

3) Changes in vacancy rate by industry

Table 7 gives estimates of changes in the job vacancy rate and standard error of this estimate for the total and for industry groups. The nine industry groups used in Tables 1, 3, and 5 (denoted by $k = 1,2,3,\ldots,9$) are again analyzed there.

To estimate the absolute value of the difference in the vacancy rate for industry groups, the vacancy rate of the firms in industry k on survey r_1 and on survey r_2 had to be calculated. These estimates are based on the 393 firms which responded to all three surveys. Formula (7) is used to estimate the job vacancy rate. Then the desired estimate of change can be represented by:

$$\left| 100\left(\frac{V}{V+E}\right)^{kr_1} - 100\left(\frac{V}{V+E}\right)^{kr_2} \right| .$$

The variance of this estimate is:

$$\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{kr_1} - 100\left(\frac{V}{V+E}\right)^{kr_2}\right] = \mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{kr_1}\right] + \mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{kr_2}\right]$$

$$- 2\,\mathrm{Cov}\left[100\left(\frac{V}{V+E}\right)^{kr_1}, 100\left(\frac{V}{V+E}\right)^{kr_2}\right] , \qquad (24)$$

where $\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{kr_1}\right]$ and $\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{kr_2}\right]$ are estimated by formula (8).

The covariance term is estimated by the following first approximation formula (see Kish, p. 210):

$$\mathrm{Cov}\left[100\left(\frac{V}{V+E}\right)^{kr_1},\ 100\left(\frac{V}{V+E}\right)^{kr_2}\right] = \frac{10{,}000\ V^{kr_1}\ V^{kr_2}}{(V+E)^{kr_1}\,(V+E)^{kr_2}}\left\{\frac{\mathrm{Cov}\ (V^{kr_1},\ V^{kr_2})}{V^{kr_1}\ V^{kr_2}}\right.$$

$$\left. + \frac{\mathrm{Cov}\left[(V+E)^{kr_1},(V+E)^{kr_2}\right]}{(V+E)^{kr_1}\ (V+E)^{kr_2}} - \frac{\mathrm{Cov}\left[V^{kr_1},(V+E)^{kr_2}\right]}{V^{kr_1}\,(V+E)^{kr_2}} - \frac{\mathrm{Cov}\left[V^{kr_2},(V+E)^{kr_1}\right]}{V^{kr_2}\,(V+E)^{kr_1}}\right\}, \qquad (25)$$

where Cov $(V^{kr_1},\ V^{kr_2})$ is estimated by formula (16).

Vacancies plus employment in industry k is defined as follows:

$$(V_{ji} + E_{ji})^{kr_1} = \begin{cases} (V_{ji} + E_{ji})^{r_1} & \text{For firms in industry } \underline{k} \text{ on survey } \underline{r}_1. \\ 0 & \text{For firms not in industry } \underline{k} \text{ on survey } \underline{r}_1. \end{cases}$$

$$(V_{ji} + E_{ji})^{kr_2} = \begin{cases} (V_{ji} + E_{ji})^{r_2} & \text{For firms in industry } \underline{k} \text{ on survey } \underline{r}_2. \\ 0 & \text{For firms not in industry } \underline{k} \text{ on survey } \underline{r}_2. \end{cases}$$

Then the covariance is:

$$\mathrm{Cov}\left[(V+E)^{kr_1},(V+E)^{kr_2}\right] = \sum_{j=1}^{9}\left[\frac{N_j^2}{n_j}(1-f_j)\right]\frac{\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{kr_1}(V_{ji}+E_{ji})^{kr_2} - \frac{\left[\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{kr_1}\right]\left[\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{kr_2}\right]}{n_j}}{n_j - 1}, \qquad (26)$$

and

$$\mathrm{Cov}\left[V^{kr_1},(V+E)^{kr_2}\right] = \sum_{j=1}^{9}\left[\frac{N_j^2}{n_j}(1-f_j)\right]\frac{\sum_{i=1}^{n_j}V_{ji}^{kr_1}(V_{ji}+E_{ji})^{kr_2} - \frac{\left[\sum_{i=1}^{n_j}V_{ji}^{kr_1}\right]\left[\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{kr_2}\right]}{n_j}}{n_j - 1}. \qquad (27)$$

Cov $\left[V^{kr_2},(V+E)^{kr_1}\right]$ is computed with a formula similar to (27), but using $V_{ji}^{kr_2}$ instead of $V_{ji}^{kr_1}$, and $(V_{ji} + E_{ji})^{kr_1}$ instead of $(V_{ji} + E_{ji})^{kr_2}$.

The difference of the total vacancy rate is:

$$\left|\ 100\left(\frac{V}{V+E}\right)^{r_1} - 100\left(\frac{V}{V+E}\right)^{r_2}\ \right|,$$

Table 7. - **Changes in the Job Vacancy Rate, Standard Error and Coefficient of Variation, by Industry Group**

Survey Period and Industry Group	Absolute Value of Change in Job Vacancy Rate (1)	Standard Error of Change in Vacancy Rate (2)	Coefficient of Variation of Change in Vacancy Rate (2)÷(1) (3)
February - May			
Durable manufacturing	0.68	0.16	0.23
Nondurable manufacturing	0.90	0.32	0.35
Construction	2.54	2.42	0.95
Public utilities and transportation	1.51	2.83	1.87
Trade, retail and wholesale	0.55	0.54	0.98
Finance, insurance, and real estate	0.60	0.37	0.62
Services, excluding education	0.12	0.71	5.95
Education	2.63	0.26	0.10
Government	0.16	0.42	2.61
Total	0.06	0.22	3.67
May - August			
Durable manufacturing	0.003	0.20	67.33
Nondurable manufacturing	1.15	0.26	0.23
Construction	1.70	1.39	0.82
Public utilities and transportation	0.78	0.56	0.72
Trade, retail and wholesale	0.12	0.44	3.69
Finance, insurance, and real estate	0.65	0.66	1.02
Services, excluding education	0.18	0.61	3.39
Education	1.64	0.08	0.05
Government	0.27	0.30	1.09
Total	0.22	0.16	0.73
February - August			
Durable manufacturing	0.68	0.22	0.32
Nondurable manufacturing	2.05	0.46	0.23
Construction	4.24	2.61	0.62
Public utilities and transportation	2.29	2.75	1.20
Trade, retail and wholesale	0.43	0.46	1.07
Finance, insurance, and real estate	1.25	0.72	0.57
Services, excluding education	0.06	0.84	13.98
Education	4.27	0.22	0.05
Government	0.11	0.54	4.88
Total	0.16	0.24	1.50

Source: Based on the 393 units that responded to all three NICB surveys.

where the vacancy rate is estimated by formula (11). The variance of this estimate is:

$$\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{r_1} - 100\left(\frac{V}{V+E}\right)^{r_2}\right] = \mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{r_1}\right] + \mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{r_2}\right]$$

$$- 2\ \mathrm{Cov}\left[100\left(\frac{V}{V+E}\right)^{r_1}, 100\left(\frac{V}{V+E}\right)^{r_2}\right] . \qquad (28)$$

Estimates of $\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{r_1}\right]$ and of $\mathrm{Var}\left[100\left(\frac{V}{V+E}\right)^{r_2}\right]$ are derived from formula (12),

and

$$\mathrm{Cov}\left[100\left(\frac{V}{V+E}\right)^{r_1}, 100\left(\frac{V}{V+E}\right)^{r_2}\right] = \frac{10{,}000\ V^{r_1} V^{r_2}}{(V+E)^{r_1}(V+E)^{r_2}}\left\{\frac{\mathrm{Cov}\ (V^{r_1}, V^{r_2})}{V^{r_1}\ V^{r_2}}\right.$$

$$\left. + \frac{\mathrm{Cov}\left[(V+E)^{r_1}, (V+E)^{r_2}\right]}{(V+E)^{r_1}\ (V+E)^{r_2}} - \frac{\mathrm{Cov}\left[V^{r_1}, (V+E)^{r_2}\right]}{V^{r_1}\ (V+E)^{r_2}} - \frac{\mathrm{Cov}\left[V^{r_2}, (V+E)^{r_1}\right]}{V^{r_2}\ (V+E)^{r_1}}\right\} , \qquad (29)$$

where $\mathrm{Cov}\ (V^{r_1}, V^{r_2})$ is estimated by formula (18),

$$\mathrm{Cov}\left[(V+E)^{r_1}, (V+E)^{r_2}\right] =$$

$$= \sum_{j=1}^{9}\left[\frac{N_j^2}{n_j}(1-f_j)\right] \frac{\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{r_1}(V_{ji}+E_{ji})^{r_2} - \frac{\left[\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{r_1}\right]\left[\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{r_2}\right]}{n_j}}{n_j - 1} , \qquad (30)$$

and

$$\mathrm{Cov}\left[V^{r_1}, (V+E)^{r_2}\right] = \sum_{j=1}^{9}\left[\frac{N_j^2}{n_j}(1-f_j)\right] \frac{\sum_{i=1}^{n_j} V_{ji}^{r_1}(V_{ji}+E_{ji})^{r_2} - \frac{\left(\sum_{i=1}^{n_j} V_{ji}^{r_1}\right)\left[\sum_{i=1}^{n_j}(V_{ji}+E_{ji})^{r_2}\right]}{n_j}}{n_j - 1} . \qquad (31)$$

Likewise, $\mathrm{Cov}\left[V^{r_2}, (V+E)^{r_1}\right]$ is computed by a formula similar to (31) where $V_{ji}^{r_1}$ is substituted by $V_{ji}^{r_2}$ and $(V_{ji} + E_{ji})^{r_2}$ is substituted by $(V_{ji} + E_{ji})^{r_1}$.

The estimates of changes of the job vacancy rate obtained are too small to be measured accurately with the NICB sample, since the coefficient of variation is greater than 1 for many industry groups. For the two three-month changes the coefficients of variation of the total are 0.71 and 3.63, and for the one six-month change it is 1.48.

4) Changes in vacancy rate by size

Table 8 presents the changes in vacancy rates for the employment size groups used in Table 6 ($s = 1,2,\ldots,5$). The absolute value of the difference in the vacancy rate is:

$$\left| 100\left(\frac{V}{V+E}\right)^{sr_1} - 100\left(\frac{V}{V+E}\right)^{sr_2} \right| , \tag{32}$$

where

$$100\left(\frac{V}{V+E}\right)^{sr_1} = \frac{100 \sum_{j=1}^{9} \frac{1}{f_j} \sum_{i=1}^{n_j} V_{ji}^{sr_1}}{\sum_{j=1}^{9} \frac{1}{f_j} \sum_{i=1}^{n_j} \left(V_{ji} + E_{ji}\right)^{sr_1}} , \tag{33}$$

and

$$(V_{ji} + E_{ji})^{sr_1} = \begin{cases} (V_{ji} + E_{ji})^{r_1} & \text{Vacancies plus employment in survey } r_1 \text{ of firm } i \text{ in size stratum } j \text{ on the selection date, and in employment size group } s. \\ 0 & \text{For firms not in size group } s \text{ on survey } r_1. \end{cases}$$

The vacancy rate for size s in survey r_2 is defined in a similar fashion, but with reference to survey r_2.

The variance of the difference of the vacancy rate is estimated by a formula similar to (24), but with reference to size groups s ($s = 1,2,3,4,5$) instead of industry groups k. The estimated variance is a first approximation.

In Table 8 it should be noted that except for the firms with 250 employees or more in all three surveys, the estimated standard error is of the same order of magnitude as the absolute value of the difference of the vacancy rate.

Table 8. **Changes in the Job Vacancy Rate, Standard Error,** and Coefficient of Variation, by Employment Size

Survey Period and Employment Size	Absolute Value of Change in Job Vacancy Rate (1)	Standard Error of Change in Vacancy Rate (2)	Coefficient of Variation of Change in Vacancy Rate (2)÷(1) (3)
February - May			
Firms that remain in same employment size			
0-9	1.19	1.40	1.18
10-49	0.97	0.59	0.61
50-249	0.12	0.92	7.67
250 and over	0.12	0.04	0.33
Firms that change employment size	0.79	1.35	1.71
Total	0.06	0.22	3.67
May - August			
Firms that remain in same employment size			
0-9	0.26	0.75	2.88
10-49	0.13	0.67	5.15
50-249	0.39	0.56	1.44
250 and over	0.19	0.04	0.21
Firms that change employment size	0.66	1.06	1.61
Total	0.22	0.16	0.73
February - August			
Firms that remain in same employment size			
0-9	1.45	1.35	0.93
10-49	1.10	0.96	0.87
50-249	0.27	0.83	3.07
250 and over	0.07	0.001	0.01
Firms that change employment size	1.45	1.80	1.24
Total	0.16	0.24	1.50

Source: Based on the 393 units that responded to all three NICB surveys.

Summary

General formulas were developed for the computation of estimates of total job vacancies, job vacancy rates, and changes in these quantities and their **variances.** Data from the NICB surveys were employed to illustrate the use of the general formulas.

The estimates of the standard error of the total vary from survey to survey. The largest estimate of the variability will be used to summarize the results, since it represents the most conservative estimate of reliability.

The coefficient of variation of estimated total job vacancies in February 1965 (Survey 1) is 0.07. The analyst may consider that total job vacancies should be estimated with a coefficient of variation of at most 0.05. This target was not achieved in the results of the 1965 NICB surveys. The variability of the estimates for industry groups and employment size of firms is important, since these estimates can be used as a guide to determine the best sample design. Three industries with a large standard error were found to be durable manufacturing, construction, and trade (retail and wholesale). However, these results may in part reflect the distribution of employment size of firms by industry. In the 1965 NICB surveys all firms with employment size of 250 or more on the selection date were included in the sample with certainty. These firms, therefore, did not contribute to the estimate of the standard error. The greater the percentage of large firms (more than 250 employees) in an industry, the smaller the estimated standard error of the industry. In the analysis of the standard error of firms classified by employment size this phenomenon is observed clearly. Firms with 1,000 employees or more have zero standard error. Job vacancies of firms with 250 to 999 employees have a relatively small standard error, which results from firms that had fewer than 250 employees on the selection date, but had more than 250 employees on the survey date. The employment size of firms with the largest standard error was that of those with fewer than 10 employees on the survey date.

To be able to measure accurately very small changes in job vacancies, very large samples are needed. When a sample is being designed, the order of magnitude of the changes that represent an important economic fact should be determined and the sample designed accordingly. For example, an area like Monroe County had a labor demand of about 276,000 in 1965. The standard error of three-month changes in vacancies of 577 and 449 job vacancies (Table 5) would measure changes of 0.4% to 0.3% of labor demand with 95% confidence. The observed change in vacancies was 829 between the dates of the February and May surveys, and 208 between the May and the August surveys. The former represented about 0.3% of labor demand (E+V), while the latter was less than 0.1% of labor demand. A much larger sample than the one used would be needed to measure such small changes accurately.

Chapter IV: A PROPOSED OPTIMAL SAMPLE DESIGN

For an optimal stratified sample design the sample sizes in each stratum are chosen so as to minimize the cost for a given variance, or, alternatively, to minimize the variance for a given cost. The application of these criteria to job vacancy surveys is developed in this chapter and illustrated by references to the Rochester area survey data. A reminder, however, is in order. A sample survey for job vacancy statistics has as its objective the obtaining of estimates of the total number of job vacancies and of job vacancy rates at certain points in time, as well as of estimates of changes in both between survey periods. An optimal sample design for one type of these statistics may or may not be optimal for others.

A) Selection of Strata for Optimal Allocation

In Chapter II, three considerations for choosing strata were given: average size, variability, and cost. Significant differences in average number and in variability of the number of job vacancies not offset by proportional differences in costs make stratification worthwhile. Specifically, it is the ratio of the standard deviation (of number of vacancies, of vacancy rate, of change in number, or of change in rate) to the square root of the cost per firm that is examined in choosing strata, together with the average value of the vacancy measure.

Owing to the nature of the lists from which job vacancy samples are drawn, the possible stratifying characteristics are restricted to industry and size of firm. We have chosen the following employment size groups for analysis, using the considerations mentioned above and the estimates of variability and cost in Tables 2 and 9.

Size Group	Number of Employees of Firm (on selection date)
1	7 or fewer (or of unknown size)
2	8 to 49
3	50 to 249
4	250 to 499
5	500 to 999
6	1,000 to 2,499
7	2,500 or more

As pointed out in Chapter III, the variability of vacancies in durable manufacturing, construction, and trade is higher than in the other industries examined (Table 1). The cost estimates (Table 9) do not offset this difference, so we have chosen the following industry classification.

Industry Group	Industry of Firm
A	Durable manufacturing, construction, and trade
B	Nondurable manufacturing, public utilities, finance, services, and government

These classifications produce 14 strata, as each size group is divided into two industry groups.

B) Methods for Obtaining an Optimal Sample Design

The purpose of designing optimal samples is the utilization of the resources available in the most efficient manner. Two alternative approaches are used to select optimal samples:

(a) To minimize the cost when the variance of the sample estimates is chosen in advance.

(b) To minimize the variance of the sample estimates when the cost is chosen in advance.

Regardless of the criterion used, the optimal sample size in a stratum is*:

$$n_g' = n \frac{\dfrac{N_g S_g}{\sqrt{c_g}}}{\sum_{g=1}^{14} \left(\dfrac{N_g S_g}{\sqrt{c_g}} \right)} \tag{1}$$

where n_g' : Number of firms in stratum g for optimal allocation.

n : Total number of firms in the sample. (n will be determined to satisfy either a predetermined variance or a fixed given cost.)

N_g : Number of firms in the population in stratum g.

c_g : Cost per firm in stratum g.

S_g : Element standard error in stratum g.

To determine the total sample size, the two criteria for obtaining an optimal sample must be considered separately. If a sample is designed to meet a given total cost,

$$C = \sum_{g=1}^{14} c_g n_g' , \tag{2}$$

by substituting the optimal value for n_g' from formula (1) in this cost function, the equation for the total sample size, n, would become the following:

$$n = \frac{C \sum_{g=1}^{14} \left(\dfrac{N_g S_g}{\sqrt{c_g}} \right)}{\sum_{g=1}^{14} (N_g S_g \sqrt{c_g})} . \tag{3}$$

---

* Cochran, p. 96.

Similarly, if a sample is designed to give a specified variance, V, then n_g' from formula (1) must be substituted in the following formula:

$$V = \sum_{g=1}^{14} N_g(N_g - n_g') \frac{S_g^2}{n_g'} . \qquad (4)$$

In this case the formula for the total sample size would result as shown:

$$n = \frac{\left(\sum_{g=1}^{14} N_g S_g \sqrt{c_g}\right)\left(\sum_{g=1}^{14} \frac{N_g S_g}{\sqrt{c_g}}\right)}{V + \sum_{g=1}^{14} N_g S_g^2} . \qquad (5)$$

To compute an optimal sample size, formula (1) also requires estimates of c_g, the cost per firm in stratum g. The total cost* of an interview survey for collecting job vacancy data is subdivided into costs assumed to be constant for all firms and costs considered proportionate to the number of vacancies in the firms. The NICB survey requested firms to describe requirements for each vacancy under active recruitment on the survey reference date. Vacancy specifications were coded according to the Dictionary of Occupational Titles.** This required repeated contacts with the firm by trained personnel to obtain complete data. Data for firms with a large number of vacancies or vague specifications were consequently more expensive to compile than those for small firms or for those with well defined requirements. The survey cost per firm in a given stratum was estimated as the sum of two components: (a) an average per firm of those costs that do not vary among firms, and (b) a cost proportionate to the average number of vacancies in each stratum.

---

* The estimate of total cost is based on the data presented in Measuring Job Vacancies, Table 5.1, p. 77.

** U.S. Bureau of Employment Security, Dictionary of Occupational Titles, Vols. I and II, Second Edition, 1949.

The data on cost per firm used for the optimal design of the sample are given in Table 9, columns (11) and (12). The range of cost per firm varies between $4.66 for firms in Size 1, Industry Group B, and $317.94 for firms in Size 7, Industry Group A. The estimated cost per firm is used in formula (1) to compute the optimal sample design.

Formula (1) also requires estimates of S_g, the element standard error of the strata selected for the optimal allocation. Since the NICB sample included all firms with 250 or more employees on the selection date, the population value of the element standard error can be obtained in the strata corresponding to these firms:

$$S_g = \sqrt{\frac{\sum_{i=1}^{N_g} V_{gi}^2 - \left(\sum_{i=1}^{N_g} V_{gi}\right)^2 / N_g}{N_g - 1}} \qquad (6)$$

where V_{gi} : Number of vacancies of firm i in stratum g.

For the strata corresponding to firms with fewer than 250 employees, two or more strata originally included in the sample were collapsed to arrive at an optimal stratum. If two strata, for example, denoted by g1 and g2, are collapsed to form stratum g, an unbiased and consistent estimate of the element variance can be taken as*:

---

* The proof that formula (7) is unbiased and consistent is given in Appendix A.

Table 9. Element Standard Error for Total Job Vacancies and for Changes in Total Job Vacancies, and Estimated Total Survey Cost Per Firm

Size of Reporting Unit (Number of Employees)	Element Standard Error of Vacancies (a)						Element Standard Error of Changes in Vacancies				Cost Per Firm	
	February		May		August		February-May		May-August			
	Industry Group (b)		Industry Group		Industry Group		Industry Group		Industry Group		Industry Group	
	A	B	A	B	A	B	A	B	A	B	A	B
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0 - 7	0.44	0.38	0.48	0.14	0.71	0.17	0.53	0.41	0.39	0.22	$ 5.28	$ 4.66
8 - 49	2.70	0.68	1.78	1.36	2.10	1.40	2.47	1.45	1.74	1.09	8.04	6.80
50 - 249	4.68	4.26	6.89	5.62	4.32	5.31	6.12	5.13	6.72	5.81	15.75	16.83
250 - 499	8.49	17.70	15.02	12.35	10.50	35.29	12.53	19.07	12.64	38.60	16.96	19.17
500 - 999	11.35	35.26	11.07	21.17	15.19	25.74	13.60	37.78	9.69	23.48	22.59	43.37
1,000 - 2,499	9.47	32.22	44.31	23.45	40.16	19.42	47.05	15.05	15.75	15.06	62.85	67.14
2,500 or more	290.56	159.22	334.50	65.09	284.25	69.85	157.97	136.34	109.51	45.89	317.94	199.39

(a) The unit of the element standard error is a job vacancy.

(b) Industry Group A includes: Firms in construction, durable manufacturing, and trade.
Industry Group B includes: Firms in nondurable manufacturing, public utilities, finance, service, and government.

Source: Columns 1-6, 11, and 12 are based on data from all units. Columns 7-10 are based on data from the 393 units that responded to all three NICB surveys.

$$s_g^2 = \frac{1}{N_g}\left\{\frac{N_{g1}}{n_{g1}}\sum_{i=1}^{n_{g1}} V_{gi}^2 + \frac{N_{g2}}{n_{g2}}\sum_{i=1}^{n_{g2}} V_{gi}^2\right\}$$

$$- \frac{1}{N_g(N_g-1)}\left\{\frac{N_{g1}}{n_{g1}}\left(\frac{N_{g1}-1}{n_{g1}-1}\right)\left[\left(\sum_{i=1}^{n_{g1}} V_{gi}\right)^2 - \sum_{i=1}^{n_{g1}} V_{gi}^2\right] + \right.$$

$$+ \frac{N_{g2}}{n_{g2}}\left(\frac{N_{g2}-1}{n_{g2}-1}\right)\left[\left(\sum_{i=1}^{n_{g2}} V_{gi}\right)^2 - \sum_{i=1}^{n_{g2}} V_{gi}^2\right]$$

$$\left. + 2\,\frac{N_{g1}}{n_{g1}}\left(\frac{N_{g2}}{n_{g2}}\right)\left(\sum_{i=1}^{n_{g1}} V_{gi}\right)\left(\sum_{i=1}^{n_{g2}} V_{gi}\right)\right\} \qquad (7)$$

where N_{g1} : Number of firms in the population in stratum g1.

N_{g2} : Number of firms in the population in stratum g2.

N_g : $N_{g1} + N_{g2}$

n_{g1} : Number of firms in the sample in stratum g1.

n_{g2} : Number of firms in the sample in stratum g2.

Formula (7) assumes that a simple random sample is chosen from each stratum separately, then later combined to form one stratum. This assumption is an approximation to the sample chosen for the NICB surveys, since within each employment-size stratum a systematic sample of firms ordered by industry was selected. The element standard error for the estimation of total job vacancies is shown in Table 9, columns (1) - (6).

To arrive at an estimate of the optimal sample design for changes in total number of vacancies in formulas (6) and (7) given above, a substitution may be made for the number of vacancies in a given survey (V_{gi}) by the difference between vacancies in any two survey periods, V_{gi}^{12} where:

$$V_{gi}^{12} = V_{gi}^{1} - V_{gi}^{2} \tag{8}$$

and V_{gi}^{1} : Number of vacancies in Survey Period 1 of firm i in stratum g.

V_{gi}^{2} : Number of vacancies in Survey Period 2 of firm i in stratum g.

The element standard error for the estimation of changes in total job vacancies is also given in Table 9, columns (7) - (10).

A first approximation formula to be used for the estimation of the element standard error for optimal allocation corresponding to the job vacancy rate $\left(\frac{100\,V}{V+E}\right)$ is as follows*:

$$\left(S_g'\right)^2 = \left[\frac{100}{(V+E)}\right]^2 \left\{ S_{V_g}^2 + \left[\frac{V}{(V+E)}\right]^2 S_{(V+E)_g}^2 - 2\left[\frac{V}{(V+E)}\right] S_{[V,(V+E)]_g} \right\} \tag{9}$$

where V : Estimated total vacancies.

$(V+E)$: Estimated total vacancies plus employment.

$S_{V_g}^2$: Element variance for vacancies in stratum g.

$S_{(V+E)_g}^2$: Element variance for vacancies plus employment in stratum g.

$S_{[V,(V+E)]_g}$: Element covariance between vacancies and vacancies plus employment in stratum g.

In the strata corresponding to firms with 250 employees or more, where the sampling fraction is taken as one, the element covariance is:

---

* Cochran, p. 175.

$$S_{[V,(V+E)]_g} = \frac{\sum_{i=1}^{N_g} V_{gi}(V+E)_{gi} - \frac{\left(\sum_{i=1}^{N_g} V_{gi}\right)\left[\sum_{i=1}^{N_g} (V+E)_{gi}\right]}{N_g}}{N_g - 1} \tag{10}$$

where $(V+E)_{gi}$: Vacancies plus employment of firm i in stratum g.

The optimal strata of firms with fewer than 250 employees were formed by combining two or more of the originally selected strata. When two or more strata are to be combined, an unbiased and consistent estimate of the element covariance is*:

$$\begin{aligned} S_{[V,(V+E)]} = {} & \frac{1}{N_g}\left[\frac{N_{g1}}{n_{g1}} \sum_{i=1}^{n_{g1}} V_{gi}(V+E)_{gi} + \frac{N_{g2}}{n_{g2}} \sum_{i=1}^{n_{g2}} V_{gi}(V+E)_{gi}\right] \\ & - \frac{1}{N_g(N_g-1)}\left\{\frac{N_{g1}}{n_{g1}}\left(\frac{N_{g1}-1}{n_{g1}-1}\right)\left[\left(\sum_{i=1}^{n_{g1}} V_{gi}\right)\left(\sum_{i=1}^{n_{g1}} (V+E)_{gi}\right) - \sum_{i=1}^{n_{g1}} V_{gi}(V+E)_{gi}\right]\right. \\ & + \frac{N_{g2}}{n_{g2}}\left(\frac{N_{g2}-1}{n_{g2}-1}\right)\left[\left(\sum_{i=1}^{n_{g2}} V_{gi}\right)\left(\sum_{i=1}^{n_{g2}} (V+E)_{gi}\right) - \sum_{i=1}^{n_{g2}} V_{gi}(V+E)_{gi}\right] \\ & \left. + \frac{N_{g1}}{n_{g1}}\left(\frac{N_{g2}}{n_{g2}}\right)\left[\left(\sum_{i=1}^{n_{g1}} V_{gi}\right)\left(\sum_{i=1}^{n_{g2}} (V+E)_{gi}\right) + \left(\sum_{i=1}^{n_{g2}} V_{gi}\right)\left(\sum_{i=1}^{n_{g1}} (V+E)_{gi}\right)\right]\right\} \end{aligned} \tag{11}$$

The estimate S_g' of the element standard error for the job vacancy rate is given in Table 10, columns (1) - (6). To obtain optimal sample sizes, S_g' should be substituted into formula (1).

---

* The proof that formula (11) is unbiased and consistent is given in Appendix A.

Table 10. Element Standard Error for the Total Job Vacancy Rate and for Changes in the Total Job Vacancy Rate

Size of Reporting Unit (Number of Employees)	Element Standard Error of Vacancy Rates(a)						Element Standard Error of Changes in Vacancy Rates					
	February		May		August		February-May		May-August		February-August	
	Industry Group(b)		Industry Group		Industry Group		Industry Group		Industry Group		Industry Group	
	A	B	A	B	A	B	A	B	A	B	A	B
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0 - 7	0.000154	0.000137	0.000160	0.000097	0.000235	0.000082	0.000198	0.000172	0.000136	0.000085	0.000170	0.000160
8 - 49	0.000941	0.000262	0.000654	0.000428	0.000697	0.000476	0.000956	0.000506	0.000605	0.000394	0.001093	0.000582
50 - 249	0.001707	0.001546	0.002330	0.001838	0.001477	0.001747	0.002043	0.001849	0.002328	0.001981	0.001137	0.002123
250 - 499	0.002446	0.006695	0.004856	0.003988	0.003351	0.012452	0.004369	0.006903	0.004490	0.013098	0.003100	0.014354
500 - 999	0.005527	0.012802	0.004936	0.007402	0.005994	0.008027	0.004878	0.013921	0.003339	0.007984	0.005339	0.015224
1,000 - 2,499	0.004570	0.013106	0.014154	0.009782	0.011038	0.008837	0.016163	0.006043	0.005892	0.005710	0.012799	0.009707
2,500 or more	0.047180	0.059130	0.099930	0.017297	0.094470	0.012478	0.055760	0.050660	0.039110	0.015450	0.059470	0.064110

(a) The unit of the element standard error is the vacancy rate (in per cent).
(b) For definition of industry groups see Table 9, footnote (b).

Source: Columns 1-6 are based on the data from all units. Columns 7-12 are based on data from the 393 units that responded to all three NICB surveys.

To calculate the optimal sample allocation for changes in the job vacancy rate, an estimate is required of the element variance for changes in the job vacancy rate. This estimate is based on the firms for which data from all three survey periods are available. The estimate of the variance for changes in the job vacancy rate is:

$$(S_g'^2)_1 + (S_g'^2)_2 - 2\, S'_{[g_1, g_2]} \qquad (12)$$

where $(S_g'^2)_1$ and $(S_g'^2)_2$ are determined by using formula (9) and represent estimates of the element variance in Survey Periods 1 and 2.

The element covariance between the two survey periods is computed by using the following first approximation*:

$$S'_{[g_1, g_2]} = \frac{100}{(V+E)_1} \left[\frac{100}{(V+E)_2}\right] \Bigg\{ S_{[V_{g1}, V_{g2}]} + \frac{V_1}{(V+E)_1}\left(\frac{V_2}{(V+E)_2}\right) \times$$

$$S_{[(V+E)_{g1}, (V+E)_{g2}]} - \frac{V_1}{(V+E)_1} S_{[V_{g2}, (V+E)_{g1}]}$$

$$- \frac{V_2}{(V+E)_2} S_{[V_{g1}, (V+E)_{g2}]} \Bigg\} \qquad (13)$$

where V_1	:	Total estimated vacancies in Survey 1.
V_2	:	Total estimated vacancies in Survey 2.
$(V+E)_1$	:	Total estimated vacancies plus employment in Survey 1.
$(V+E)_2$	:	Total estimated vacancies plus employment in Survey 2.

---

* Kish, p. 210.

$S_{(V_{g1},V_{g2})}$: Element covariance between vacancies in Surveys 1 and 2 in stratum g.

$S_{[(V+E)_{g1},(V+E)_{g2}]}$: Element covariance between vacancies plus employment in Surveys 1 and 2 in stratum g.

$S_{[V_{g2},(V+E)_{g1}]}$: Element covariance between vacancies in Survey 2 and vacancies plus employment in Survey 1 in stratum g.

$S_{[V_{g1},(V+E)_{g2}]}$: Element covariance between vacancies in Survey 1 and vacancies plus employment in Survey 2 in stratum g.

The element covariances for strata of firms with 250 or more employees are estimated using formulas similar to (10). For firms with fewer than 250 employees — in which the optimal strata were derived by collapsing two or more original strata — the element covariance can be calculated by formulas similar to (11). In both cases, appropriate variables must be used. For example, to estimate the element covariance between vacancies in Survey 1 and Survey 2, the variables V_{gi}^{1} and V_{gi}^{2} should be used. The estimates of the element standard error of changes in the vacancy rate are given in Table 10, columns (7) - (12).

In the next section all these formulas are applied to data obtained in the NICB surveys of Monroe County in order to illustrate optimal designs for job vacancy surveys.

C) Optimal Sample Design for Estimating Total Number of Job Vacancies

Table 11 presents optimal sample sizes for total job vacancies, using the following constraints: (a) a total cost of $9,500 (which was the estimated cost in 1965 prices of a continuing sample survey for obtaining job vacancy data of the size used by the NICB*), or (b) coefficients of variation of 0.025, 0.05, or 0.10. Sample designs for specified costs are given in columns (1) and (2) of Table 11. The total sample size is approximately 600 employers. Table 12, which gives optimal sampling rates, shows that in the strata for employers with fewer than 50 employees, a lower sampling ratio is optimal for those in Industry Group B than for the small employers in Industry Group A (employers in construction, durable manufacturing, and trade industries). Employers with 50 to 249 employees have similar sampling ratios for all industries.

If the optimal sampling fraction in a stratum is larger than 0.5, including all firms is considered advisable. This criterion was discussed by Leslie Kish as follows:

> For computing means, aggregates, and their variances the formulas of Section 3.3 are applied. In the variance formula...it may be noted again that increasing all the allocations n_h proportionally by the factor k decreases the overall variance by the same factor — if changes in $\bar{f}_h$ are negligible.
>
> But optimum allocation is precisely the method that may lead to situations where the values $(1 - f_h)$ in some strata are too large to be negligible. Moreover, sometimes in one or more of the extreme strata the formula...may point to values of f_h close to 1 or even over 1. In such cases, all elements of these strata should be taken into the sample. The extreme strata, if sampled completely, do not contribute to the variance of the combined estimate.**
>
> . . . .
>
> The optimum formulas...may result in the impossible allocation $n_h > N_h$ and $f_h > 1$. This can occur when a rather large sample is needed

---

* Based on data presented in Measuring Job Vacancies, p. 77.

** Kish, p. 94.

Table 11. Optimal Sample Size for Estimates of Total Job Vacancies

Survey Period and Size of Reporting Unit (Number of Employees)	Number of Firms in Optimal Sample for Given Cost of $9,500		Number of Firms in Optimal Sample for Coefficient of Variation of:						Per Cent Distribution of Firms in Optimal Sample Design (c)	
	Industry Group (a)		0.025		0.05		0.10		Industry Group	
			Industry Group		Industry Group		Industry Group			
	A	B	A	B	A	B	A	B	A	B
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
February										
0 - 7	122	129	357	377	133	140	47	50	19.7	20.8
8 - 49	191	32	557	94	207	35	74	12	30.7	5.2
50 - 249	32	29	95	84	35	31	13	11	5.2	4.6
250 - 499	8	19	22	31	8	21	3	7	1.2	3.1
500 - 999	6	17	17	21	7	19	2	7	1.0	2.8
1,000 - 2,499	2(b)	5	3	9	2(b)	6	2(b)	2	0.1	0.9
2,500 or more	9	4	9	4	9	4	9	3	3.6	1.1
Total	605		1,680		657		242		100.0	
May										
0 - 7	156	55	335	118	135	47	46	16	25.8	9.1
8 - 49	138	70	296	151	119	61	41	21	22.8	11.6
50 - 249	54	42	117	91	47	37	16	13	9.0	7.0
250 - 499	15	15	24	31	13	13	4	4	2.5	2.4
500 - 999	7	11	14	21	6	10	2	3	1.1	1.9
1,000 - 2,499	5	4	5	9	4	4	2(b)	2(b)	0.8	0.7
2,500 or more	9	3	9	4	9	3	8	2(b)	4.7	0.5
Total	584		1,225		508		180		100.0	
August										
0 - 7	203	59	540	157	203	59	75	22	31.8	9.2
8 - 49	148	61	394	163	148	61	55	23	23.2	9.6
50 - 249	30	37	80	97	30	37	11	13	4.7	5.7
250 - 499	9	31	23	31	9	31	3	14	1.4	6.0
500 - 999	8	13	17	21	8	13	3	5	1.3	2.0
1,000 - 2,499	4	3	5	9	4	3	2(b)	2(b)	0.6	0.5
2,500 or more	9	3	9	4	9	3	8	2(b)	3.5	0.5
Total	618		1,550		618		238		100.0	

(a) For definition of industry groups see Table 9, footnote (b).
(b) The number of firms in this stratum was increased to two, the minimum number necessary to compute an estimate of the variance.
(c) **Owing to rounding, percentage distributions do not necessarily add to total.**

Source: Based on data from all units that responded to the NICB surveys.

Table 12. Optimal Sampling Ratios of Firms for Estimates of Total Job Vacancies

Survey Period and Size of Reporting Unit (Number of Employees)	Optimum for Given Cost $9,500 Industry Group (a) A (1)	B (2)	Optimum for Coefficient of Variation of: 0.025 Industry Group A (3)	B (4)	0.05 Industry Group A (5)	B (6)	0.10 Industry Group A (7)	B (8)
February								
0 - 7	1:34	1:38	1:12	1:13	1:31	1:35	1:89	1:98
8 - 49	1:7	1:25	3:7	1:8	1:6	1:23	1:17	1:66
50 - 249	2:11	1:6	1:2	6:13	1:5	1:6	1:14	1:16
250 - 499	1:3	19:31	22:24	1:1	1:3	21:31	1:8	7:31
500 - 999	6:17	17:21	1:1	1:1	7:17	19:21	2:17	1:3
1,000 - 2,499	2:5	5:9	3:5	1:1	2:5	2:3	2:5	2:9
2,500 or more	1:1	1:1	1:1	1:1	1:1	1:1	1:1	3:4
May								
0 - 7	1:28	1:91	1:13	1:43	1:32	1:107	1:95	1:314
8 - 49	1:9	1:11	1:4	1:5	1:11	1:13	1:31	1:38
50 - 249	3:10	2:9	5:8	1:2	1:4	1:5	1:12	1:14
250 - 499	5:8	15:31	1:1	1:1	13:24	13:31	1:6	4:31
500 - 999	7:17	11:21	14:17	1:1	6:17	10:21	2:17	1:7
1,000 - 2,499	1:1	4:9	1:1	1:1	4:5	4:9	2:5	2:9
2,500 or more	1:1	3:4	1:1	1:1	1:1	3:4	8:9	1:2
August								
0 - 7	1:21	1:85	1:8	1:32	1:21	1:85	1:57	1:228
8 - 49	1:9	1:12	1:3	2:9	1:9	1:12	1:23	1:32
50 - 249	1:6	1:5	4:9	1:2	1:6	1:5	1:16	1:14
250 - 499	9:23	1:1	1:1	1:1	9:23	1:1	3:23	14:31
500 - 999	8:17	13:21	1:1	1:1	8:17	13:21	3:17	5:21
1,000 - 2,499	4:5	1:3	1:1	1:1	4:5	1:3	2:5	2:9
2,500 or more	1:1	3:4	1:1	1:1	1:1	3:4	8:9	1:2

(a) For definition of industry groups see Table 9, footnote (b).

Source: Based on data from all units that responded to the NICB surveys.

> from a skewed population of establishments or inventories. In the extreme strata the values of $\underline{S}_h$ may be so large that the formula leads to the impossible preliminary result. This solution lies in the mathematical region where the contribution of the extreme strata to the sum of variances...would be negative. The first practical step is to make these variances zero by making $\underline{f}_h = 1$ and $\underline{n}_h = \underline{N}_h$. In practice we may decide to take this step for any stratum where $\underline{f}_h > 0.5$, let us say.*

The use of the criterion that all sampling units are included in each stratum with an optimal sampling fraction of 0.5 or more results in a sample which comprises all firms with 1,000 or more employees and those with 250 to 999 employees in Industry Group B. One out of every two firms with 250-999 employees in Industry Group A should be selected for the design of this optimal sample.

An estimate of the variance of vacancies is needed to compute the optimal sample size from equation (13). Since the coefficient of variation (ratio of standard error to the estimated statistic) has been predetermined, the variance can be estimated as follows:

$$\text{Variance} = (\text{Vacancies} \times \text{Coefficient of variation})^2$$

The average sample size necessary to obtain estimates of total vacancies with a coefficient of variation of 0.05 is estimated at about 600 firms (Table 11). This is similar to the sample size corresponding to a given total cost of $9,500. To obtain survey results with a coefficient of variation of 0.025, a sample of about 1,500 firms is needed, while if it is sufficient to estimate total job vacancies with a coefficient of variation of 0.10, a sample of 220 firms would suffice.

In Table 13 the average optimal sample size for total job vacancies, as well as the largest sample size per stratum, is given. If the largest sample size is selected for each stratum, the variance of the resulting sample will be no greater than that specified for the optimal

---

* Ibid., p. 144.

Table 13. Average Optimal Sample Size and Largest Optimal Sample Size

for Total Job Vacancies

Stratum		Optimum for Given Cost $9,500		Optimum for Coefficient of Variation of					
				0.025		0.05		0.10	
Number of Employees	Industry Group	Average Size (a)	Largest Size (b)	Average Size	Largest Size	Average Size	Largest Size	Average Size	Largest Size
		NUMBER OF FIRMS							
0 - 7	A	160	203	411	540	157	203	56	75
	B	81	129	217	377	82	140	29	50
8 - 49	A	159	191	416	557	158	207	57	74
	B	54	70	136	163	52	61	19	23
50 - 249	A	39	54	97	117	37	47	13	16
	B	36	42	91	97	35	37	12	13
250 - 499	A	11	15	23	24	10	13	3	4
	B	22	31	31	31	22	31	8	14
500 - 999	A	7	8	16	17	7	8	2	3
	B	14	17	21	21	14	19	5	7
1,000 - 2,499	A	4	5	4	5	3	4	2	2
	B	4	5	9	9	4	6	2	2
2,500 or more	A	9	9	9	9	9	9	8	9
	B	3	4	4	4	3	4	2	3
Total		603	783	1,485	1,971	593	789	218	295

(a) Arithmetic mean of the three optimal sample sizes given in Table 11.
(b) Largest of the three optimal sample sizes in each stratum given in Table 11.

Source: Based on data from all units that responded to the NICB surveys.

sample design, but the cost will be increased. For example, for the coefficient of variation of 0.05, the sample including the largest size per stratum results in a total sample size of 789 firms. This sample includes over 100 more firms than the largest optimal sample size for a coefficient of variation of 0.05, which is 657 firms for the data of Survey 1.

If, on the other hand, we assume that the results of the three surveys differ only on account of sampling variations and take an average of the three survey results, the total sample size for a coefficient of variation of 0.05, including all firms in strata where the sampling fraction is larger than 0.5, is as follows:

Table 14. Selected Sample Design for Estimating Total Job Vacancies to Obtain a Coefficient of Variation of 0.05

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a]		Sampling Ratio of Firms in Industry Group	
	A	B	A	B
0-7	157	82	1:27	1:61
8-49	158	52	1:8	1:15
50-249	37	35	1:5	1:5
250-499	10	31[b]	5:12	1:1
500-999	7	21[b]	7:17	1:1
1,000-2,499	5[b]	4	1:1	4:9
2,500 or more	9	4[b]	1:1	1:1
Total	612			

(a) For definition of industry groups see Table 9, footnote (b).
(b) The average optimal sample sizes were increased in these strata to include all firms, since the optimal sampling fraction was at least 0.5.

Source: Based on Table 13.

The cost of one survey which follows the optimal sample design given in Table 14 is approximately $10,100, and theoretically the sample estimates would result in a coefficient of variation for total job vacancies of 0.05.

The sample used in the NICB surveys of Monroe County in 1965 included with certainty all firms with 250 employees or more on the selection date. The optimal design suggested above included with certainty firms with 250 or more employees except employers with 250-999 employees in Industry Group A, and those with 1,000-2,499 employees in Industry Group B, where one out of every two firms was selected. The characteristics of the firms which should be selected with certainty usually vary according to the size of the area being covered and the distribution of firms by employment size and industry.

D) Optimal Sample Design for Estimating Changes in Number of Total Job Vacancies

The criterion to be used for the design of the optimal sample to measure changes in job vacancies which occurred between survey periods is based on the samples detecting changes of 0.5% of total labor demand (employment plus vacancies) with 95% confidence. In 1965 the average labor demand for Monroe County was 276,000 jobs. The number of vacancies which represents one-half of one per cent would therefore be 1,380. If the sample estimate of change has a standard error of 690 vacancies, then changes in vacancies of one-half of one per cent are detectable with 95% confidence.

Optimal sample designs for a standard error of change of 690 vacancies for three-month periods are given in Table 15. The optimal sample for changes which occurred between Survey 1 and Survey 2 requires 381 firms, while one based on the changes which occurred between Survey 2 and Survey 3 requires only 278 firms. If all firms in strata where the sampling fraction is 0.5 or larger are included with certainty, an optimal sample design for estimating changes in total job vacancies with a standard error of 690 vacancies would be as follows:

Table 15. Optimal Sample Size for Estimates of Changes in Total Job Vacancies

Size of Reporting Unit (Number of Employees)	Optimum for Standard Error of Change of 690 Job Vacancies					
	February-May		May-August		Average Sample Size (b)	
	Industry Group (a)		Industry Group		Industry Group	
	A (1)	B (2)	A (3)	B (4)	A (5)	B (6)
	NUMBER OF FIRMS					
0 - 7	76	79	53	41	64	60
8 - 49	99	36	65	25	82	30
50 - 249	24	20	25	21	24	20
250 - 499	6	12	6	22	6	17
500 - 999	4	10	3	6	4	8
1,000 - 2,499	3	2(c)	2(c)	2(c)	2	2
2,500 or more	7	3	5	2(c)	6	2
Total	381		278		327	

(a) For definition of industry groups see Table 9, footnote (b).
(b) Arithmetic mean of the two optimal designs given in columns 1-4 of this Table.
(c) See Table 11, footnote (b).

Source: Based on data from the 393 units that responded to all three NICB surveys.

Table 16. Selected Sample Design for Estimating Changes in Total Job Vacancies Which Measure 0.5% of (E+V) with 95% Confidence

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a]		Sampling Ratio of Firms in Industry Group	
	A	B	A	B
0-7	64	60	1:59	1:82
8-49	82	30	1:16	1:25
50-249	24	20	1:8	1:9
250-499	6	31[b]	6:23	1:1
500-999	4	8	4:17	8:21
1,000-2,499	5[b]	2	1:1	2:9
2,500 or more	9[b]	4[b]	1:1	1:1
Total	349			

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 14, footnote (b).

Source: Based on Table 15.

The total sample size described in the table above is smaller than the sample size used for the NICB surveys. The distribution by strata of the 393 employers that responded to the three NICB surveys is the following:

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a]	
	A	B
0-7	44	47
8-49	63	40
50-249	38	42
250-499	23	31
500-999	17	21
1,000-2,499	5	9
2,500 or more	9	4
Total	393	

(a) For definition of industry groups see Table 9, footnote (b).

Source: NICB surveys.

However, if the sample sizes in the 14 strata are analyzed separately, it will be noted that the sample is not reduced in all strata. In fact, the optimal sample size is larger than the NICB sample for employers with 7 or fewer employees in both industry groups and for those with 8-49 employees in Industry Group A. Therefore, the distribution of the optimal sample for the estimation of changes in job vacancies differs from the distribution of the sample used for the NICB surveys.

The optimal sampling ratios for estimates of three-month changes in job vacancies are given in Table 17. For the strata corresponding to the smaller firms, those with fewer than 8 employees and those with 8-49 employees, the optimal sampling ratio is larger for firms in Industry Group A than for those in Industry Group B. However, for firms with 250-499 employees, a sampling ratio of one out of every four firms in Industry Group A is optimal, while approximately one out of every two firms should be included in Industry Group B. For firms with 2,500 or more employees, the optimal sampling ratio is at least one of every two firms. In this case all firms should be included with certainty. Besides the arguments given for the adoption of this criterion in Section C, an open-end stratum may include a giant employer. If this employer were omitted from the sample, the sample estimates might be very inaccurate.

Table 16, which gives an average sample size based on the optimal sample designs computed for the two three-month survey periods in Table 15, assumes that the differences between the two estimates are due to sampling variation only. If significant differences exist between the two periods, the optimal sample design to be chosen should select from each stratum the largest number of sampling units. This has been done in Table 18, below.

Table 17. Optimal Sampling Ratios of Firms for Estimates of Changes in Total Job Vacancies

Size of Reporting Unit (Number of Employees)	Optimum for Standard Error of Change of 690 Job Vacancies			
	February-May		May-August	
	Industry Group (a)		Industry Group	
	A	B	A	B
0 - 7	1:50	1:62	1:71	1:120
8 - 49	1:13	1:20	1:20	1:29
50 - 249	1:8	1:9	1:7	1:9
250 - 499	1:4	12:31	1:4	22:31
500 - 999	4:17	10:21	3:17	2:7
1,000 - 2,499	3:5	2:9	2:5	2:9
2,500 or more	7:9	3:4	5:9	1:2

(a) For definition of industry groups see Table 9, footnote (b).

Source: Based on data from the 393 units that responded to all three NICB surveys.

Table 18. Largest Optimal Sample Size for Each Stratum to Estimate Three-month Changes in Vacancies of 0.5% of (E+V) with 95% Confidence

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a] A	B	Sampling Ratio of Firms in Industry Group A	B
0-7	76	79	1:50	1:62
8-49	99	36	1:13	1:20
50-249	25	21	1:7	1:9
250-499	6	31[b]	6:23	1:1
500-999	4	10	4:17	10:21
1,000-2,499	5[b]	2	1:1	2:9
2,500 or more	9[b]	4[b]	1:1	1:1
Total	407			

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 14, footnote (b).

Source: Based on Table 15.

The total sample size includes 58 more employers than the average optimal sample presented in Table 16. The sample given in Table 18 is selected so as to provide estimates of three-month changes in vacancies which measure 0.5% of employment plus vacancies in Monroe County with 95% confidence. The sample includes with certainty all employers with 2,500 employees or more, also those with 250-499 employees in Industry Group B, and those with 1,000-2,499 employees in Industry Group A.

E) Optimal Sample Design for Estimating the Job Vacancy Rate

The job vacancy rate (100V/V+E) is a counterpart measure to the unemployment rate. In other words, the unemployment rate measures the number of people seeking jobs (those who on the reference date were not holding a job) relative to the labor force (employed plus unemployed seeking employment) in the given area; the vacancy rate, on the other hand, measures the number of jobs which employers are seeking to fill on the reference date, relative to employment plus vacancies (a measure of labor demand) for the given area. The job vacancy rate converts the total number of job vacancies to a relative measure, thus permitting comparison of job situations in areas of differing size.

Table 19 shows optimal sample designs to estimate job vacancy rates with a coefficient of variation of 0.05 and 0.10. To obtain a coefficient of variation of 0.05, the optimal sample size computed varies between 535 and 623 employers. However, for a coefficient of variation of 0.10 the total sample size is reduced to between 201 and 262 employers. Table 20 gives the optimal sampling ratios to estimate the job vacancy rate with a coefficient of variation of 0.05 and with a coefficient of variation of 0.10. In this table, it should be noted that the sampling ratios for employers with fewer than 50 employees are larger for those in Industry Group A than for those in Industry Group B. For example, optimal sampling ratios in Survey 3 for employers with fewer than 8 employees were 1:22 for those in Industry Group A, while they were 1:60 for those in Industry Group B. For employers with 50-249 employees the optimal sampling ratios were similar for both industry groups.

In Table 21 an average optimal sample design to estimate job vacancy rates with a coefficient of variation of 0.05 is presented, the sample sizes being increased to include all units in the strata where the average optimal sampling fraction was at least 0.5:

Table 19. Optimal Sample Size for
Estimates of the Total Job Vacancy Rate

Survey Period and Size of Reporting Unit (Number of Employees)	Number of Firms in Optimal Sample for Coefficient of Variation of				Per Cent Distribution of Firms in Optimal Sample Design (c)	
	0.05		0.10			
	Industry Group (a)		Industry Group		Industry Group	
	A	B	A	B	A	B
February						
0 - 7	119	131	51	56	19.7	21.5
8 - 49	184	34	79	15	30.4	5.6
50 - 249	33	29	14	12	5.4	4.8
250 - 499	6	20	3	9	1.0	3.3
500 - 999	8	17	4	7	1.4	2.9
1,000 - 2,499	2(b)	6	2 (b)	3	0.2	1.0
2,500 or more	9	4	4	3	1.7	1.2
Total	602		262		100.0	
May						
0 - 7	126	92	46	33	23.1	16.8
8 - 49	124	54	45	20	22.7	9.9
50 - 249	45	34	16	12	8.2	6.2
250 - 499	12	12	4	4	2.2	2.1
500 - 999	7	10	3	4	1.3	1.8
1,000 - 2,499	4	4	2(b)	2	0.7	0.8
2,500 or more	9	2	8	2(b)	3.8	0.4
Total	535		201		100.0	
August						
0 - 7	193	84	71	31	30.0	13.0
8 - 49	143	60	52	22	22.2	9.4
50 - 249	30	35	11	13	4.6	5.4
250 - 499	8	31	3	14	1.3	6.1
500 - 999	10	11	3	4	1.5	1.8
1,000 - 2,499	3	4	2(b)	2	0.5	0.7
2,500 or more	9	2	8	2(b)	3.3	0.2
Total	623		238		100.0	

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 11, footnote (b).
(c) **Owing to rounding, percentage distributions do not necessarily add to total.**

Source: Based on data from all units that responded to the NICB surveys.

Table 20. Optimal Sampling Ratios of Firms for Estimates of the **Total Job Vacancy Rate**

Survey Period and Size of Reporting Unit (Number of Employees)	Optimum for Coefficient of Variation of			
	0.05		0.10	
	Industry Group(a)		Industry Group	
	A	B	A	B
February				
0 - 7	1:35	1:38	1:82	1:88
8 - 49	1:7	1:23	1:16	1:53
50 - 249	2:11	1:6	1:13	1:15
250 - 499	1:4	20:31	1:8	9:31
500 - 999	8:17	17:21	4:17	1:3
1,000 - 2,499	2:5	2:3	2:5	1:3
2,500 or more	1:1	1:1	4:9	3:4
May				
0 - 7	1:35	1:55	1:95	1:152
8 - 49	1:10	1:15	1:29	1:40
50 - 249	1:4	1:5	1:12	1:15
250 - 499	1:2	12:31	1:6	4:31
500 - 999	7:17	10:21	3:17	4:21
1,000 - 2,499	4:5	4:9	2:5	2:9
2,500 or more	1:1	1:2	8:9	1:2
August				
0 - 7	1:22	1:60	1:60	1:162
8 - 49	1:9	1:12	1:25	1:33
50 - 249	1:6	1:5	1:16	1:14
250 - 499	8:23	1:1	3:23	14:31
500 - 999	10:17	11:21	3:17	4:21
1,000 - 2,499	3:5	4:9	2:5	2:9
2,500 or more	1:1	1:2	8:9	1:2

(a) For definition of industry groups see table 9, footnote (b).

Source: Based on data from all units that responded to the NICB surveys.

Table 21. Selected Sample Design for Estimating the Job Vacancy Rate with a Coefficient of Variation of 0.05

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a]		Sampling Ratio of Firms in Industry Group	
	A	B	A	B
0-7	146	102	1:29	1:49
8-49	150	49	1:9	1:16
50-249	36	33	1:5	2:11
250-499	9	31[b]	3:8	1:1
500-999	8	21[b]	8:17	1:1
1,000-2,499	5[b]	9[b]	1:1	1:1
2,500 or more	9	4[b]	1:1	1:1
Total	612			

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 14, footnote (b).

Source: Based on Table 26.

The optimal sample size in Table 21 is 612 employers. The sample includes with certainty all employers with at least 1,000 employees, and employers in Industry Group B with 250-999 employees. For employers with 250-999 employees in Industry Group A, a sampling ratio of approximately 1:2 would be adequate. All employers with 50-249 employees should be selected with a sampling ratio of 1:5. The sampling ratios corresponding to the strata of employers with fewer than 50 employees are smaller for employers in Industry Group B than for those in Industry Group A within each employment size category. The cost of one survey using the sample distribution suggested in Table 21 would be approximately $10,300, on the basis of the costs per stratum given in Table 9.

Alternatively, instead of selecting an average of the three optimal samples for vacancy rates, the largest sample size per stratum can be selected. Table 22 below corresponds to this latter method.

Table 22. Largest Optimal Sample Size for Each Stratum to Estimate the Job Vacancy Rate with a Coefficient of Variation of 0.05

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a] A	Number of Firms in Industry Group[a] B	Sampling Ratio of Firms in Industry Group A	Sampling Ratio of Firms in Industry Group B
0-7	193	131	1:22	1:38
8-49	184	60	1:7	1:13
50-249	45	35	1:4	1:5
250-499	24[b]	31	1:1	1:1
500-999	17[b]	21[b]	1:1	1:1
1,000-2,499	5[b]	9[b]	1:1	1:1
2,500 or more	9	4	1:1	1:1
Total	768			

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 14, footnote (b).

Source: Based on Table 19.

The average optimal sample size for estimating the job vacancy rate is 612 employers; the sample obtained by selecting the largest number of sampling units per stratum requires 768 employers. This latter sample includes with certainty all employers with 250 employees or more. The cost of the sample given in Table 21 was approximately $10,300, while the sample in Table 22 would cost about $11,700. In Section G of this chapter the various possible samples will be compared to determine the most desirable alternative in terms of reliability of results in relation to cost.

F) Optimal Sample Design for Estimating Changes in the Job Vacancy Rate

Changes in the job vacancy rate suggest a tightening or loosening of the labor market. One of the objectives of the job vacancy surveys is to detect an economically significant change with reasonable confidence. In Table 23 optimal sample designs for achieving a standard error of 0.25 are given. These samples are designed to detect with 95% confidence changes

Table 23. Optimal Sample Size for Estimates of Changes in the **Total Job Vacancy Rate**

Size of Reporting Unit (Number of Employees)	Optimum for Standard Error of Changes in Vacancy Rates of 0.25					
	February-May		May-August		February-August	
	Industry Group(a)		Industry Group		Industry Group	
	A	B	A	B	A	B
	NUMBER OF FIRMS					
0 - 7	80	94	50	43	66	85
8 - 49	107	35	62	25	118	39
50 - 249	23	20	23	19	12	22
250 - 499	6	12	6	21	4	24
500 - 999	4	11	3	6	4	11
1,000 - 2,499	3	2	2(b)	2(b)	2	3
2,500 or more	7	4	4	2(b)	7	4
Total	408		268		401	

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 11, footnote (b).

Source: Based on the 393 units that responded to all three NICB surveys.

in the job vacancy rate of 0.5% or more. For example, the job vacancy rate was approximately 3% in Monroe County in 1965. The sample designs given in Table 23 would detect with 95% confidence changes in the vacancy rate of either from 3.0% to 3.5% or from 3.0% to 2.5%. If it were required to detect smaller changes, the sample size would need to be increased.

The estimated optimal sample size for three-month changes in the vacancy rate with a standard error of 0.25 is 268 and 408 employers. The average optimal sample size for three-month changes (Table 26) is 337 employers. Table 25 gives optimal sampling ratios for changes in the two three-month periods (February-May and May-August), and also for the six-month period (February-August).

In Table 24 below, an average optimal sample for estimating three-month changes in the job vacancy rate of 0.5 with 95% confidence is given. Here, in the strata where the computed sampling fraction was at least 0.5, the sample size was increased to include all employers.

Table 24. Selected Sample Design for Estimating Three-month Changes of 0.5 or More in the Job Vacancy Rate with 95% Confidence

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a]		Sampling Ratio of Firms in Industry Group	
	A	B	A	B
0-7	65	68	1:58	1:72
8-49	84	30	1:15	1:25
50-249	23	20	1:8	1:9
250-499	6	31[b]	6:23	1:1
500-999	4	8	4:17	8:21
1,000-2,499	2	2	2:5	2:9
2,500 or more	9[b]	4[b]	1:1	1:1
Total	356			

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 14, footnote (b).

Source: Based on Table 26.

Table 25. Optimal Sampling Ratios of Firms for Estimates of Changes in **the Total Job Vacancy Rate**

Survey Period and Size of Reporting Unit (Number of Employees)	Optimum for Standard Error of Changes in Vacancy Rate of 0.25 Industry Group(a) A	B
February-May		
0 - 7	1:47	1:52
8 - 49	1:12	1:21
50 - 249	1:8	1:9
250 - 499	6:23	12:31
500 - 999	4:17	11:21
1,000 - 2,499	3:5	2:9
2,500 or more	7:9	1:1
May-August		
0 - 7	1:76	1:114
8 - 49	1:21	1:29
50 - 249	1:8	1:9
250 - 499	6:23	21:31
500 - 999	3:17	2:7
1,000 - 2,499	2:5	2:9
2,500 or more	4:9	1:2
February-August		
0 - 7	1:57	1:58
8 - 49	1:11	1:19
50 - 249	1:15	1:8
250 - 499	4:23	24:31
500 - 999	4:17	11:21
1,000 - 2,499	2:5	1:3
2,500 or more	7:9	1:1

(a) For definition of industry groups see Table 9, footnote (b).

Source: Based on the 393 units that responded to all three NICB surveys.

Table 26. Average Optimal Sample Size for Vacancy Rates **and Three-month Changes in Vacancy Rates**

Size of Reporting Unit (Number of Employees)	Average Optimal Sample **Size^a for Estimating Vacancy** Rates for Coefficient of Variation of:				Average Optimal Sample **Size^b** of Three-month Changes for Standard Error of Changes in Vacancy Rate of 0.25	
	0.05		0.10			
	Industry Group(c)		Industry Group		Industry Group	
	A	B	A	B	A	B
	(1)	(2)	(3)	(4)	(5)	(6)
	NUMBER OF FIRMS					
0 - 7	146	102	56	40	65	68
8 - 49	150	49	59	19	84	30
50 - 249	36	33	14	12	23	20
250 - 499	9	21	3	9	6	16
500 - 999	8	13	3	5	4	8
1,000 - 2,499	3	5	2	2	2	2
2,500 or more	9	3	7	2	6	3
Total	587		233		337	

(a) Arithmetic mean of the three optimal sample sizes given in Table 19.
(b) Arithmetic mean of the February-May and May-August optimal designs given in Table 23.
(c) For definition of industry groups see Table 9, footnote (b).

Source: Columns 1-4 are based on data from all units. Columns 5 and 6 are based on data from the 393 units that responded to all three NICB surveys.

The sample size given in Table 24 suggests including 356 employers. The sample includes with certainty employers with 2,500 employees or more and employers with 250-499 employees in Industry Group B. For strata corresponding to employers with fewer than 50 employees, the pattern of a smaller optimal sampling ratio for those in Industry Group B than for those in Industry Group A is similar to the result found in the optimal sample design for estimating the job vacancy rate.

The total cost of the sample given in Table 24 is approximately $7,300. However, an average taken of two sample sizes assumes that the difference between them is due to sampling variation only. If, instead, the largest sample size is selected from each stratum, the standard of reliability established for designing the optimal sample is ensured. Table 27 below corresponds to this latter criterion:

Table 27. Largest Optimal Sample Size of Each Stratum for Estimating Three-month Changes of 0.5 or More in the Job Vacancy Rate with 95% Confidence

Size of Reporting Unit (Number of Employees)	Number of Firms in Industry Group[a]		Sampling Ratio of Firms in Industry Group	
	A	B	A	B
0-7	80	94	1:47	1:52
8-49	107	35	1:12	1:21
50-249	23	20	1:8	1:9
250-499	6	31[b]	6:23	1:1
500-999	4	21[b]	4:17	1:1
1,000-2,499	5[b]	2	1:1	2:9
2,500 or more	9[b]	4	1:1	1:1
Total	441			

(a) For definition of industry groups see Table 9, footnote (b).
(b) See Table 14, footnote (b).

Source: Based on Table 23.

Four hundred and forty-one employers are required for the sample given in Table 27. The cost of this sample is approximately $8,500, which is $1,200 more than that of the average sample given in Table 24, but $1,000 less than the estimated cost of the NICB surveys. Both the sample discussed in Table 24 and the one in Table 27 are designed to measure three-month changes in the vacancy rate of 0.5 with 95% confidence.

G) Comparison of Sample Designs

Job vacancy surveys should yield estimates not only of the number of vacancies and of the job vacancy rate, but also of the changes in vacancies and vacancy rates between survey periods. In Sections C-F, optimal sample designs to estimate each of the four statistics mentioned above have been discussed. The different optimal allocations are compared below.

Table 28. Per Cent Distribution of Employer Units in Optimal Sample Designs for Monroe County Compared with the NICB Surveys of Monroe County

Stratum		Average Per Cent Distribution[b]					Per Cent Distribution of NICB Sample Restricted to Units that Responded to All Surveys
		Optimal Sample to Estimate			Optimal Sample to Estimate Changes in Three-month Periods of		
Number of Employees	Industry Group[a]	Job Vacancies	Vacancy Rate	NICB Sample	Job Vacancies	Vacancy Rate	
		(1)	(2)	(3)	(4)	(5)	(6)
0-7	A	25.8	24.3	12.7	19.5	19.3	11.2
	B	13.0	17.1	11.9	17.9	19.6	12.0
8-49	A	25.6	25.1	15.6	24.9	24.9	16.0
	B	8.8	8.3	10.2	9.3	9.0	10.2
50-249	A	6.3	6.1	9.5	7.7	7.2	9.7
	B	5.8	5.5	10.5	6.4	6.1	10.7
250-499	A	1.7	1.5	5.8	1.9	1.8	5.9
	B	3.8	3.8	7.7	5.6	5.4	7.9
500-999	A	1.1	1.4	4.2	1.1	1.0	4.3
	B	2.2	2.2	5.2	2.5	2.4	5.3
1,000-2,499	A	0.5	0.5	1.2	0.5	0.5	1.3
	B	0.7	0.8	2.2	0.4	0.5	2.3
2,500 or more	A	3.9	2.9	2.2	1.7	1.7	2.3
	B	0.7	0.6	1.0	0.6	0.6	1.0
Total		100.0	100.0	100.0	100.0	100.0	100.0

(a) For definition of industry groups see Table 9, footnote (b).
(b) Owing to rounding, percentage distributions do not necessarily add to total.

Sources: See following page.

The optimal percentage distributions for estimating job vacancies and the job vacancy rate on the survey reference date are similar, except for the stratum corresponding to employers with fewer than 8 employees in Industry Group B (Table 28). The distribution of the NICB sample, originally designed to estimate total job vacancies, included all firms with 250 or more employees (column (3) of Table 28),while the optimal samples for both job vacancies and job vacancy rates call for only partial coverage in this size range. The following tabulation summarizes the percentages which the larger employers represent in the three samples.

Sample	Percentage of Total Sample Represented by Employers with 250 or More Employees
Optimal for	
Job Vacancies	14.6
Job Vacancy Rates	13.6
NICB	29.5

Source: Table 28.

The principal reason for these differences is the wide variation in cost per employer, shown in Table 9. In designing the NICB sample, these costs were erroneously assumed to be the same for all sizes of firm, owing to the absence of information.

Sources for Table H on preceding page:

Column (1) is based on Table 11.

Column (2) is based on Table 19.

Column (3) is the per cent distribution of the arithmetic mean of the number of employers that responded to the NICB surveys.

Column (4) is the average optimal per cent distribution for estimating changes in vacancies, based on the 393 employers that responded to all three NICB surveys.

Column (5) is the average optimal per cent distribution for estimating changes in the vacancy rate, based on the 393 employers that responded to all three NICB surveys.

Column (6) is based on the 393 employers that responded to all three NICB surveys.

The two distributions computed to estimate three-month changes, given in columns (4) and (5), Table 28, are almost identical, but they differ from the samples given in columns (1) and (2) of Table 28, especially in the following strata:

(a) Employers with 0-7 employees, Industry Group A.

(b) Employers with 250-499 employees, Industry Group B.

(c) Employers with at least 2,500 employees, Industry Group A.

In Table 29 below, the cost and total sample size of the optimal samples described in the text tables of Sections C to F of this chapter are given. In columns (1) and (2) average samples are given; in columns (3) and (4) the total sample size was formed by choosing the largest sample size for each stratum. All employers in the stratum were included when the samples were computed if the sampling fraction was at least 0.5.

Table 29. Sample Size and Cost of Job Vacancy Surveys

Criteria for Optimal Design	Average Optimal Sample		Largest Sample Size per Stratum	
	Number of Employers (1)	Cost (2)	Number of Employers (3)	Cost (4)
Coefficient of Variation of 0.05:				
Job Vacancies	612	$10,081	806	$11,837
Job Vacancy Rate	612	10,324	768	11,688
Three-month Changes with 95% Confidence:				
Job Vacancies of 0.5% of (E+V)[a]	349	7,436	407	7,885
Job Vacancy Rates of 0.5%	356	7,291	441	8,462

(a) E+V, employment plus vacancies represents total labor demand.

Sources: Tables 9, 13, 14, 16, 18, 21, 22, 24, and 27.

To obtain a coefficient of variation of 0.05, the average sample size for both estimates of job vacancies and of the job vacancy rate is 612 employers. Yet the cost of the two samples differs slightly because the composition of the samples by stratum varies (see Table 28), and the estimates of cost per firm in the separate strata are also different (see Table 9).

The criteria specified for estimates of changes in job vacancies and in job vacancy rates can be satisfied with smaller and, therefore, less costly samples. If a larger sample is selected and the sample size in each stratum is larger than that needed to estimate changes of 0.5% with 95% confidence, then all requirements are met. Table 30, below, gives the strata with optimal percentages for estimates of change that are larger than for estimates of totals.

Table 30. Strata in Which Estimates for Changes Exceed Those for Totals: Percentages of Total Samples and Their Ratios

Stratum		Percentage to Estimate			Percentage to Estimate		
Number of Employees	Industry Group[a]	Job Vacancies	3-month Changes in Vacancies	Ratio (2)÷(1)	Vacancy Rates	3-month Changes in Vacancy Rates	Ratio (5)÷(4)
		(1)	(2)	(3)	(4)	(5)	(6)
0-7	B	13.0	17.9	1.38	17.1	19.6	1.15
8-49	B	8.8	9.3	1.06	8.3	9.0	1.08
50-249	A	6.3	7.7	1.22	6.1	7.2	1.18
50-249	B	5.8	6.4	1.10	5.5	6.1	1.11
250-499	A	1.7	1.9	1.12	1.5	1.8	1.20
250-499	B	3.8	5.6	1.47	3.8	5.4	1.42
500-999	B	2.2	2.5	1.14	2.2	2.4	1.09

(a) For definition of industry groups see Table 9, footnote (b).

Source: Table 28.

In all other strata, if a total sample size larger than the optimal for estimates of change is selected, estimates of changes between survey periods with the desired accuracy will necessarily be obtained. In the case of the strata listed in Table 30, if a sample size meets the required precision of the estimates of total job vacancies, it is possible that, even when the optimal percentage for estimates of change is larger, the sample size chosen may in fact be adequate. For example, in Table 14, in the stratum with fewer than 8 employees in Industry Group B, a sample size of 82 units is required; in Table 16, which gives the average sample for estimating changes in vacancies, the sample corresponding to the same stratum comprises only 60 units. Therefore, if a sample of 612 units is chosen (including 82 units in the above mentioned stratum), the resulting estimates of changes in vacancies will also be of the necessary precision in this stratum. If the total sample size for total vacancies is 50% larger than the sample size for estimates of changes, it follows that the sample sizes in each stratum will be large enough to satisfy all criteria. This holds since the largest ratio of 1.47 occurs in the stratum of employers with 250-499 employees in Industry Group B.

However, in Table 31, it must be noted that the sample sizes for estimates of job vacancies and of the job vacancy rate are always larger by more than 50% than the estimated optimal sample sizes for changes.

Table 31. Ratio of Estimates for One Survey Period to Estimates for Three-Month Changes for: Optimal Sample Size and Cost

Estimates	Ratio	
	Sample Size (Number of Employers)	Cost
Average of Job Vacancies to Average of 3-month Changes in Vacancies	1.75	1.36
Average of Job Vacancy Rate to Average of 3-month Changes in Rates	1.72	1.42
Largest of Job Vacancies to Largest of 3-month Changes in Vacancies	1.98	1.50
Largest of Job Vacancy Rates to Largest of 3-month Changes in Rates	1.74	1.38

Source: Table 29.

Therefore, if the optimal sample sizes for estimating job vacancies or job vacancy rates with a coefficient of variation of 0.05 are chosen, the sample will also meet the requirements for estimating changes of 0.5% with 95% confidence.

In the selection of a sample which meets all the criteria established in Sections C to F, the largest sample size for each stratum should be chosen. Table 32 gives the resulting samples.

Table 32. Largest Sample Size for Each Stratum of the Designs Given in Sections C to F

Size of Reporting Unit and Cost	Average Optimal Sample in Industry Group[a] A	Average Optimal Sample in Industry Group[a] B	Largest Optimal Sample in Industry Group A	Largest Optimal Sample in Industry Group B
0-7	157	102	203	140
8-49	158	52	207	61
50-249	37	35	47	37
250-499	10	31	24	31
500-999	8	21	17	21
1,000-2,499	5	9	5	9
2,500 or more	9	4	9	4
Total	638		815	
Total Cost	$10,533		$12,040	

(a) For definition of industry groups see Table 9, footnote (b).

Sources: Tables 9, 13, 14, 16, 18, 21, 22, 24, and 27.

Although the table given above shows the largest sample size in each stratum, the sample size usually corresponds to the estimates for the total number of job vacancies (Section C); possible exceptions might occur, however, in the three strata listed in Table 33 below.

Table 33. Strata in Which Optimal Percentages Are Larger for Estimates of Vacancy Rates than for Estimates of Job Vacancies

Stratum		Optimal Percentages for Estimating	
Number of Employees	Industry Group[a]	Job Vacancies	Job Vacancy Rates
0-7	B	13.0	17.1
500-999	A	1.1	1.4
1,000-2,499	B	0.7	0.8

(a) For definition of industry groups see Table 9, footnote (b).

Source: Table 28.

If the several differing estimates result from sampling variation only, then it is sufficient to choose the average sample given in Table 32, which comprises 638 sampling units and costs approximately $10,500. This sample includes about 200 more sampling units than the NICB survey of Monroe County did, and would cost about $1,000 more. However, the desired precision of this sample is a coefficient of variation of 0.05 for estimates of job vacancies and of the job vacancy rate, while the achieved NICB precision is a coefficient of variation of 0.07.

If, on the other hand, the variations among the estimates of Survey Periods 1, 2, and 3 are considered to be due to seasonal variations or to other economically significant factors, and not exclusively to sampling variability, the largest sample should be chosen in each stratum. This sample includes 815 sampling units and costs approximately $12,000. The NICB surveys provide three quarterly observations. If significant seasonal or other variations exist, estimates for the fourth quarter would be needed to provide adequate data on which to base the optimal sample design.

If Tables 18 and 27 are compared and the largest sample size in each stratum is chosen, the following sample for the purpose of estimating three-month changes would result.

Table 34. Largest Sample Size for Each Stratum of the Designs to Estimate Three-month Changes Given in Tables 18 and 27

Size of Reporting Unit and Cost	Number of Employers in Industry Group[a] A	 B
0-7	80	94
8-49	107	36
50-249	25	21
250-499	6	31
500-999	4	21
1,000-2,499	5	2
2,500 or more	9	4
Total	445	
Total Cost	$8,517	

(a) For definition of industry groups see Table 9, footnote (b).

Sources: Tables 9, 18, and 27.

Therefore, if the survey is designed to estimate changes in vacancies of 0.5% of total labor demand (E+V) or changes of 0.5% in the job vacancy rate with 95% confidence, a sample of 445 sampling units (employers) will suffice. Such a sample would cost approximately $8,500 per survey. It is interesting to note that although this sample is larger than the sample used in the NICB surveys of Monroe County, the estimated cost per survey is $1,000 less.

Summary

We have adopted the following criteria for the selection of samples of adequate precision:

(a) A coefficient of variation not greater than 0.05 for total job vacancies. This is equivalent to a 95% confidence interval, from 0.90V' to 1.10V', for two standard errors are 10% of V', the sample estimate of vacancies

(b) A coefficient of variation not greater than 0.05 for the job vacancy rate. This is equivalent to a 95% confidence interval from 0.90R' to 1.10R' where R' is the sample estimate of the job vacancy rate

(c) Detection, with 95% confidence, of a change of 0.5 or more in the number of vacancies as a per cent of total labor demand (employment plus vacancies)

(d) Detection, with 95% confidence, of a change of 0.5 or more in the job vacancy rate

Techniques for sample selection using these criteria are developed and illustrated with data from the NICB 1965 surveys in Monroe County, New York. The conclusions listed below depend to some extent on the nature of the area studied. However, we believe that they are of general applicability to areas with other industrial structures and labor market situations.

1. Stratification by size of firm and by industry improves sampling efficiency.

2. Survey costs vary significantly by size of employer in surveys seeking detailed description of each class of job vacancy. This consideration should be taken into account to obtain an efficient sample design.

3. The criteria for measuring changes, (c) and (d) above, may be met with smaller, and therefore less costly samples than the criteria for measuring total job vacancies or the job vacancy rate, (a) and (b) above.

4. A sample designed to measure total vacancies or the vacancy rate is probably adequate to measure changes as well (if the criteria stated above are reasonable). This was true of the designs developed for Rochester, although the relative distributions among the strata differ according to the criterion selected.

Chapter V: GENERAL METHODS FOR DETERMINING STRATA AND SAMPLE SIZES

In this chapter certain general procedures for the design of optimal survey samples are analyzed, with the objective of establishing which ones would be applicable to areas with differing characteristics. A certain homogeneity of methods is advisable for vacancy surveys, especially those which cover different areas, but at the same period of time. The statistics thus derived could provide bases for estimates on population subclassifications (industries, for example).

Statistical theory recommends that certain characteristics, such as the number of establishments and the patterns of vacancies in an area, be taken into account in designing a vacancy survey. Consideration of such characteristics not only influences sample design in general, but,in particular, the choice of both strata boundaries and sample sizes as well. Optimal strata and optimal sample sizes in each area will depend to a large degree on the size of the area chosen and on its prevailing job vacancy rate.

A) The Choice of Stratification Variables

Theoretically, the best stratifying variable would be the one to be measured in the survey (see Chapter II, Section A and Chapter IV, Section A). Therefore, if employers can be stratified according to the number of vacancies they have, maximum gains are obtained from stratification. When this variable is not available, some alternatives may be employed, such as:

a) a variable highly correlated with the variable to be measured, or

b) a variable which classifies the population of establishments into such domains of study as are desirable for job vacancy statistics.

If the first criterion mentioned above is used, stratification of establishments by number of employees is presumed advantageous, since larger establishments are likely to have more job vacancies than small establishments.

The relationship between job vacancies and employment is not necessarily linear, however, inasmuch as vacancies depend on current labor shortages and on expected labor turnover, as well as on any projected expansion or curtailment of employment in firms.* Moreover, new operations have higher job vacancy rates than stabler ones, and these, in turn, have a higher rate than those scheduled to close. In the NICB surveys, the highest average immediate vacancy rate of 3.7 corresponds to those firms with 10-49 employees, and the estimated rate is lower for both employers with 50 or more employees and those with fewer than 10 employees.

In regard to b), the second alternative, the distribution of job vacancies by industry, is one of the measures to be derived from a job vacancy survey and is a characteristic of the sampling unit. Therefore, stratification by industry would improve the quality of estimates of vacancies by industry.

B) The Choice of the Number of Strata

The total variance for any survey estimate, S^2, can be divided into a portion, S_s^2, which can be reduced when a stratified sample is chosen plus another portion which is independent of stratification, S_e^2.

$$S^2 = S_s^2 + S_e^2$$

It has been shown**that the portion of the variance susceptible to reduction by the use of stratification decreases inversely with the square of the number of strata. This relationship is exact for a rectangular distribution and approximate for samples selected from skewed distributions. As the number of strata increases,the portion of the variance independent of stratification will soon dominate the estimate of total variance. To increase the number of strata beyond 6* does not, it is generally agreed, increase the accuracy of estimates in relation to the increase in cost, except when the strata correspond to subgroups for which separate estimates are desired. In the present study the sample design for Rochester, N. Y., uses seven size strata and each of these is substratified into two industry groups. This gives fourteen size-industry strata. For a larger labor area, a greater number of industry substrata might be advisable.

---

* A more precise discussion of the factors determining job vacancies can be found in John G. Myers, "Job Vacancies in the Firm and the Labor Market," The Conference Board, May, 1968.

** Cochran, p. 133.

C) The Formation of Strata in the Population

The rule in use to determine the best boundaries for size strata was suggested by Dalenius and Hodges**. It has been found both efficient and easy to apply***. Let f_y be the frequency function of establishments classified by number of employees in a given area; if the establishments have been grouped into classes of unequal width, let d_y be the number of a predetermined standardized unit in each size group. Then the cumulative of $\sqrt{f_y d_y}$ is formed and the total range is divided into equal intervals which, in turn, determine the best boundaries of the strata. The data from the NICB surveys of Rochester, N. Y., have been used here to find optimal size strata by following this rule. In the table below optimal size strata were also calculated for New York, N. Y., and Richmond, Virginia****.

Before determining the best boundaries between size strata, the very large establishments of the area were separated and grouped to form a stratum that could be included with certainty, since large establishments are likely to include a great proportion of the vacancies and the estimated variance is also very large (see discussion in Chapter IV, Section D). For all other establishments six optimal size strata were calculated, as shown below.

---

* Ibid., p. 134.

** Tore Dalenius and Joseph L. Hodges, Jr., "Minimum Variance Stratification," Journal of the American Statistical Association, March 1959, pp. 88-101.

*** William G. Cochran, "Comparison of Methods for Determining Stratum Boundaries," Bulletin of the International Statistical Institute, Vol. 38 (1961), 2, pp. 345-358.

**** The distribution of establishments by size was taken from County Business Patterns 1965, except for the distribution of establishments with 500 or more employees in New York, N. Y., which was obtained from the New York office of the Bureau of Employment Security.

Table 35. Optimal Size Strata for Rochester, N. Y., New York, N. Y., and Richmond, Virginia, in 1965

	Six Optimal Strata for Establishments with					
	Fewer than 2,500 Employees in Rochester, N. Y.		Fewer than 5,000 Employees in New York, N. Y.		Fewer than 500 Employees in Richmond, Va.	
Stratum	Number of Employees	Stratum Width	Number of Employees	Stratum Width	Number of Employees	Stratum Width
1	4 or fewer	4	5 or fewer	5	4 or fewer	4
2	5-20	16	6-21	16	5-13	9
3	21-79	59	22-73	52	14-35	22
4	80-255	176	74-249	176	36-83	48
5	256-770	515	250-960	711	84-196	113
6	771-2,499	1,729	961-4,999	4,039	197-499	303
Number of Establishments With	2,500 or more employees		5,000 or more employees		500 or more employees	
	14		34		30	
Number of Employees	247,001		3,828,271		154,914	

Sources: For distribution of establishments, see text.
For number of employees: County Business Patterns 1965.

From data in Table 35 the effect of the distribution of establishments according to number of employees on the computation of six optimal strata may be calculated. The calculations shown in the table determine the best boundaries in each case. In New York, N. Y., an area with 3.8 million employees and with 270,000 establishments in 1965, the computation of six optimal size strata has included establishments with fewer than 5,000 employees. This leaves 34 employers with at least 5,000 employees which could be included with certainty. In Richmond, Virginia, which had 150,000 employees and 8,800 establishments in 1965, there were 30 establishments with 500 or more employees; these establishments constitute the certainty strata for this area. Since the six strata were computed for establishments with fewer than 500 employees, the strata are much finer than the ones derived for the New York metropolitan statistical area.

D) $\underline{n_g}$: The Sample Size in Stratum g

In this section, optimal sample sizes for Rochester, N. Y., are determined by the procedure outlined in Chapter IV, Section B. The analysis demonstrates that sample sizes must be changed as job vacancy rates vary among areas, if a constant level of reliability is to be attained.

By way of illustration,the table below gives immediate job vacancy rates for standard metropolitan statistical areas surveyed by the Department of Labor in 1965 and 1966.

Table 36. Estimated Rate of Current Job Vacancies by Area, April 1, 1966 and April 15, 1965

SMSA	Estimated Total Current Job Vacancy Rate[(a)] April 1, 1966	April 15, 1965
Baltimore	2.0[(b)]	1.1
Birmingham	1.1	0.9
Charleston, S. C.	1.9	2.1
Charleston, W. Va.	0.7	0.9
Chicago	1.7	1.1
Hartford	3.7	Not available
Kansas City	1.1	0.6
Los Angeles	1.5	0.9
Miami	1.8	1.1
Milwaukee	1.7	1.7
Minneapolis-St. Paul	2.4	1.3
New York	1.4	1.0
Portland, Ore.	1.4	1.0
Richmond, Va.	2.2	2.0

(a) Estimated number of vacancies as a per cent of the sum of estimated current employment and vacancies.

(b) May 1966.

Source: U. S. Department of Labor, Report on 1966 Job Vacancy Surveys, May 10, 1967.

According to this table, the range of the job vacancy rate lies between 3.7 for Hartford in 1966 and 0.6 for Kansas City in 1965.

The NICB surveys of Rochester, N. Y., found a vacancy rate there in 1965 of approximately 3.0, a figure which, however, cannot be considered comparable to the estimates given in the Department of Labor surveys, since the NICB surveys estimated both immediate vacancies and those with a future starting date. The NICB survey data are used here to determine sample sizes which estimate vacancies with a coefficient of variation of 0.05 and of 0.10. This computation requires estimates of the variability of vacancies and of the cost per firm for each stratum defined.

The cut-off point for the computation of six optimal size strata necessarily varies from one area to another. In Rochester, N. Y., an area with 250,000 employees and 12,500 establishments in 1965, firms with fewer than 2,500 employees were included in the computation presented in Table 35, while in Richmond, Va., for example, the six strata were computed from establishments with fewer than 500 employees. The optimal boundaries for Rochester, N.Y., must be modified to conform with the size strata actually selected for the NICB sample. This must be done to arrive at estimates of the element variance which can be derived only for the strata originally chosen or for any combination of them.

The following size strata approximate the boundaries found in Table 35 and will be used here to compute optimal sample sizes.

Size Group	Number of Employees of Establishments (on the selection date)
1	7 or fewer (or of unknown size)
2	8 to 19
3	20 to 49
4	50 to 249
5	250 to 749
6	750 to 2,499
7	2,500 or more

Each size stratum is substratified into two industry groups to produce fourteen size-industry strata. The two industry groups used are the same as those defined in Chapter IV.

Industry Group	Industry of Establishment
A	Durable manufacturing, construction, and trade
B	Nondurable manufacturing, public utilities, finance, services, and government

For each size-industry stratum the element standard error and cost per firm are computed, and the results are shown in Table 37.

Table 37. Element Standard Errors for Total Job Vacancies and Estimated Survey Cost Per Firm for Rochester Surveys in 1965

Size of Reporting Unit (Number of Employees)	Element Standard Error for Vacancies(a)						Cost Per Firm	
	February		May		August			
	Industry Group(b)		Industry Group		Industry Group		Industry Group	
	A	B	A	B	A	B	A	B
0-7	0.44	0.38	0.48	0.14	0.71	0.17	$5.28	$4.66
8-19	1.29	0.57	1.55	1.20	1.94	1.41	7.48	5.95
20-49	4.62	0.83	2.30	1.57	2.50	1.40	8.66	7.34
50-249	4.68	4.26	6.89	5.62	4.32	5.31	15.75	16.83
250-749	10.89	22.15	13.87	16.66	11.36	31.78	18.94	24.81
750-2,499	8.11	38.89	30.79	23.00	31.97	25.49	33.58	59.86
2,500 or more	290.56	159.22	334.50	65.09	284.25	69.85	317.94	199.39

(a) The unit of the element standard error is a job vacancy. The computations are carried out using formulas 6 or 7 of Chapter IV.

(b) For definition of industry groups see Table 9, footnote (b).

Source: Based on data from all units that responded to the NICB surveys.

These data may also be used to estimate sample sizes which would meet established levels of reliability if job vacancy rates were 2.5, 2.0, 1.5, or 1.0. However, such estimates would be valid only if the distribution of establishments, the variability of vacancies, and the cost per firm coincided with the ones for Rochester, N. Y. Nevertheless, the computation serves to point up the increased size of samples necessary to obtain the same level of reliability for estimates of a population with a smaller number of total vacancies than in Rochester in 1965.

Table 38 gives the optimal sample sizes for the three 1965 NICB surveys and the estimated cost for each sample based on the element standard error and cost per firm shown in Table 37. The variations between the estimated sample sizes for each of the three quarterly observations shown in this table may be due to seasonal or to sampling variations.

If seasonal variations in the patterns of job vacancies by stratum are significantly different, optimal samples should be chosen by selecting the largest sample size in each stratum. If it is assumed that the fourth quarter (for which no data are available) would not differ significantly from the three estimates actually obtained, a sample so chosen would insure that the estimates of vacancies meet the specified reliability requirements. For example, in Table 39 such samples are given for the vacancies of Rochester in 1965, which correspond to a job vacancy rate of 3.0, approximately.

Table 38. Optimal Sample Sizes for Estimating Total Job Vacancies in Rochester, 1965, and for a Job Vacancy Rate of 1.5

Month and Number of Employees	Number of Establishments in Optimal Sample for (a)							
	Job Vacancy Rate of Rochester, 1965 (Approximately 3.0)				Assumed Job Vacancy Rate: 1.5			
	Coefficient of Variation				Coefficient of Variation			
	0.05		0.10		0.05		0.10	
	Industry Group(b)		Industry Group		Industry Group		Industry Group	
	A	B	A	B	A	B	A	B
February								
0-7	127	134	47	49	348	367	131	138
8-19	71	19	26	7	194	53	73	20
20-49	90	15	33	5	247	40	93	15
50-249	34	30	12	11	92	82	35	31
250-749	12	31	4	11	30	44	12	32
750-2,499	4	14	2[c]	5	10	17	4	14
2,500 or more	9	4	9	3	9	4	9	4
Total	594		224		1,537		611	
Total Cost	$9,497		$5,584		$17,248		$9,641	
May								
0-7	134	47	46	16	360	127	148	52
8-19	77	37	27	13	208	98	85	40
20-49	41	25	14	9	110	68	45	28
50-249	47	37	16	13	125	98	51	40
250-749	14	21	5	7	30	44	15	23
750-2,499	12	7	4	3	16	17	14	8
2,500 or more	9	3	9	2[c]	9	4	9	3
Total	511		184		1,314		561	
Total Cost	$8,692		$5,097		$16,575		$9,233	
August								
0-7	198	58	75	22	543	158	201	58
8-19	100	40	38	15	273	109	101	40
20-49	46	23	17	9	126	64	47	24
50-249	29	36	11	14	80	98	30	36
250-749	11	42	4	16	29	44	11	43
750-2,499	13	8	5	3[c]	16	17	13	9
2,500 or more	9	3	8	2[c]	9	4	9	3
Total	616		239		1,570		625	
Total Cost	$9,557		$5,257		$17,619		$9,697	

(a) Formulas 1 and 5, Chapter IV, are used to determine optimal sample sizes.
(b) For definition of industry groups see Table 9, footnote(b).
(c) The number of establishments in this stratum was increased to two, the minimum number necessary to compute an estimate of the variance.

Source: Based on data from all units that responded to the NICB surveys.

Table 39. Largest Optimal Sample Size of Each Stratum for Estimating Total Job Vacancies Found in Rochester, with Coefficients of Variation of 0.05 and 0.10

Size of Reporting Unit (Number of Employees)	Coefficient of Variation of 0.05		Coefficient of Variation of 0.10	
	Number of Firms in Industry Group(a)		Number of Firms in Industry Group	
	A	B	A	B
0-7	198	134	75	49
8-19	100	40	38	15
20-49	90	25	33	9
50-249	47	37	16	14
250-749	14	42	5	16
750-2,499	13	14	5	5
2,500 or more	9	4	9	3
Total	767		292	
Total Cost	$11,223		$6,256	

(a) For definition of industry groups see Table 9, footnote (b).

Source: Based on Tables 37 and 38.

The sample size thus obtained in Table 39 for a coefficient of variation of 0.05 would include 150 more establishments than the largest optimal sample given in Table 38 for a job vacancy rate of 3.0. The cost would also be increased by about $1,700 to a total of $11,200. Furthermore, if an estimate of vacancies with a coefficient of variation of 0.10 is all that is required, a sample of 292 establishments would be sufficient and the cost would be lowered to approximately $6,300.

On the other hand, the assumption can be made that the three quarterly estimates given in Table 38 differ from one another due only to sampling variability, so that the sample to be used should be an average of the three quarterly observations. The results of this procedure for various job vacancy rates are shown in Table 40. For an area like Rochester a sample of 574 establishments would thus provide estimates of vacancies with a coefficient of variation of 0.05. The cost of this sample is estimated at $9,200, which is less than the estimated cost of an NICB survey for the number of establishments which were included in the

Table 40. Average Optimal Sample Sizes for Estimating Total Job Vacancies According to Specified Job Vacancy Rates

Coefficient of Variation and Number of Employees	Vacancies of Rochester, N. Y., 1965 (Job Vacancy Rate of Approximately 3.0) Industry Group(a) A	B	Assumed Job Vacancy Rate: 2.5 Industry Group A	B	Assumed Job Vacancy Rate: 2.0 Industry Group A	B	Assumed Job Vacancy Rate: 1.5 Industry Group A	B	Assumed Job Vacancy Rate: 1.0 Industry Group A	B
	NUMBER OF ESTABLISHMENTS									
Coefficient of Variation of 0.05										
0-7	153	80	201	104	279	145	417	217	684	354
8-19	83	32	108	42	150	58	225	87	368	142
20-49	59	21	77	28	107	39	161	57	249	94
50-249	37	34	48	45	67	62	99	93	156	152
250-749	12	31	16	38	22	42	30	44	30	44
750-2,499	10	10	12	13	13	15	14	17	16	17
2,500 or more	9	3	9	4	9	4	9	4	9	4
Total	574		745		1,012		1,474		2,319	
Total Cost	$9,202		$11,075		$13,380		$17,157		$23,594	
Coefficient of Variation of 0.10										
0-7	56	29	73	38	105	54	160	83	283	147
8-19	30	12	39	15	57	22	86	33	153	59
20-49	21	8	28	10	40	15	62	22	109	39
50-249	13	13	18	16	25	23	39	36	67	63
250-749	4	11	6	15	9	22	13	33	23	43
750-2,499	4	4	5	5	7	7	10	10	13	15
2,500 or more	9	2	9	2	9	3	9	3	9	4
Total	216		279		398		599		1,027	
Total Cost	$5,373		$6,025		$7,431		$9,449		$13,516	

(a) For definition of industry groups see Table 9, footnote (b).

Source: Based on all units that responded to the NICB surveys, Tables 37, and 38.

1965 samples, yet it included about 170 more firms. In other words, an optimal sample design for Rochester would include a larger number of small establishments, which have a smaller unit cost, and would cover a smaller number of large establishments (250-2,499 employees) than the NICB sample.

These illustrative computations demonstrate that, to provide the desired precision in estimates, the total sample size must necessarily be increased as the job vacancy rate declines. The size of the labor area (as measured by the number of establishments) also influences the size of the sample to be used. The data presented in Table 40 are only applicable to a labor area similar to Rochester as regards establishments, variability of vacancies, and cost per firm. Nonetheless, our data illustrates the relation of the sample size to:

a) the job vacancy rate and/or

b) the required reliability of the estimates

In synthesis, the lower the vacancy rate in an area, the larger will be the sample size needed to meet requirements of reliability, and the less stringent the reliability requirements, the smaller the size of the sample needed. Therefore, a sample of 1,000 establishments would be adequate to measure vacancies with a coefficient of variation of only 0.10 if the vacancy rate were 1.0, yet for a vacancy rate of 2.0, this same sample size would estimate vacancies with a coefficient of variation of 0.05.

Summary

The design of an optimal sample for job vacancy surveys in a specific area should take into account:

1) the number of establishments (sampling units),

2) the distribution of establishments into strata according to number of employees and industry,

3) the differences among strata in vacancy variability and in survey cost per firm, and

4) the job vacancy rate.

Tabulations have been furnished showing optimal size strata for three different areas: New York, N. Y., Rochester, N. Y., and Richmond, Va. Illustrative computations demonstrate that the size of establishments in an area will determine both the cut-off point for that stratum which comprises the largest establishments to be sampled with certainty and those optimal boundaries which are to be used in defining each size stratum. The suggested method for determining stratum boundaries can easily be applied to areas not previously surveyed, since it is based on the distribution of establishments by number of employees, such data being readily available.

The data compiled in the NICB surveys have been used to compute optimal sample sizes for a range of possible job vacancy rates. In this connection it has been shown that sample sizes for an area like Rochester quadruple as the vacancy rate diminishes from 3.0 to 1.0 in estimating total vacancies with a coefficient of variation of 0.05. For any area, as the job vacancy rate diminishes, larger samples must be used to estimate the corresponding number of vacancies with the same reliability.

A less stringent requirement of reliability for vacancies reduces sample sizes. For example, while a vacancy rate of 2.0 requires a sample of 1,000 to estimate vacancies with a coefficient of variation of 0.05, a sample of 400 would be sufficient for a coefficient of variation of 0.10.

Appendix A: PROOFS OF UNBIASED AND CONSISTENT ESTIMATES OF ELEMENT VARIANCE AND ELEMENT COVARIANCE

To prove that formula (7) is an unbiased estimate of equation (6), Chapter IV, when simple random samples are chosen separately from two strata and afterwards collapsed to form one stratum, consider the population with values:

$$\underbrace{Y_{11}\ Y_{21}\ \cdots\ Y_{N_1 1}}_{\text{Stratum 1}} \qquad \underbrace{Y_{12}\ Y_{22}\ \cdots\ Y_{N_2 2}}_{\text{Stratum 2}}$$

$N = N_1 + N_2$ is the total size of the population. From the population select a simple random sample of size n_1 from stratum 1 and a simple random sample of size n_2 from stratum 2. Denote the sample values as follows:

$$\underbrace{y_{11}\ y_{21}\ \cdots\ y_{n_1 1}}_{\text{Stratum 1}} \qquad \underbrace{y_{12}\ y_{22}\ \cdots\ y_{n_2 2}}_{\text{Stratum 2}}$$

$n = n_1 + n_2$ is the total size of the sample selected from strata 1 and 2.

The element standard error of the population, S^2, is:

$$S^2 = \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 - \frac{\left(\sum_{i=1}^{N} Y_i\right)^2}{N}\right\}$$

$$= \frac{1}{N-1}\left\{\sum_{i=1}^{N} Y_i^2 - \left(\frac{\sum_{i=1}^{N} Y_i^2 + 2\sum_{i\neq j}^{N} Y_i Y_j}{N}\right)\right\}$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{2}{N(N-1)} \sum_{\substack{i \\ i\neq j}}^{N} Y_i Y_j \,. \qquad (1)$$

Consider the following:

$$E\left[\sum_{i=1}^{n} y_i^2\right] = E\left[\sum_{i=1}^{n_1} (y_{i1})^2 + \sum_{i=1}^{n_2} (y_{i2})^2\right] .$$

The maximum number of possible samples which include n_1 elements from stratum 1 and n_2 elements from stratum 2 are:

$$\binom{N_1}{n_1} \binom{N_2}{n_2} .$$

The number of samples which include a particular elements i of stratum 1 are:

$$\binom{N_1-1}{n_1-1} \binom{N_2}{n_2} .$$

The number of samples which include a particular element i of stratum 2 are:

$$\binom{N_1}{n_1} \binom{N_2-1}{n_2-1} .$$

Therefore,

$$E\left(\sum_{i=1}^{n} y_i^2\right) = \frac{\binom{N_1-1}{n_1-1}\binom{N_2}{n_2}\sum_{i=1}^{N_1}(Y_{i1})^2 + \binom{N_1}{n_1}\binom{N_2-1}{n_2-1}\sum_{i=1}^{N_2}(Y_{i2})^2}{\binom{N_1}{n_1}\binom{N_2}{n_2}}$$

$$= \frac{n_1}{N_1}\sum_{i=1}^{N_1}(Y_{i1})^2 + \frac{n_2}{N_2}\sum_{i=1}^{N_2}(Y_{i2})^2$$

and

$$E\left[\frac{1}{N}\left(\frac{N_1}{n_1}\sum_{i=1}^{n_1}(y_{i1})^2 + \frac{N_2}{n_2}\sum_{i=1}^{n_2}(y_{i2})^2\right)\right] = \frac{1}{N}\left[\sum_{i=1}^{N_1}(Y_{i1})^2 + \sum_{i=1}^{N_2}(Y_{i2})^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \tag{2}$$

An unbiased estimate of the first term of equation (1) is:

$$\frac{1}{N}\left(\frac{N_1}{n_1}\sum_{i=1}^{n_1}(y_{i1})^2 + \frac{N_2}{n_2}\sum_{i=1}^{n_2}(y_{i2})^2\right) . \tag{3}$$

To find an unbiased estimate of the second term of equation (1) we note that: (a) The number of samples including two distinct elements i and j from stratum 1 are

$$\binom{N_1-2}{n_1-2}\binom{N_2}{n_2} .$$

(b) The number of samples including two distinct elements i and j from stratum 2 are $\binom{N_1}{n_1}\binom{N_2-2}{n_2-2}$.

(c) The number of samples including an element i from stratum 1 and an element j from stratum 2 are

$\binom{N_1-1}{n_1-1}\binom{N_2-1}{n_2-1}$.

$$E\left[\sum_{\substack{1\\ i\neq j}}^{n} y_i y_j\right]$$

$$= \frac{\binom{N_1-2}{n_1-2}\binom{N_2}{n_2}\sum\limits_{\substack{1\\ i\neq j}}^{N_1} Y_{i1}Y_{j1} + \binom{N_1}{n_1}\binom{N_2-2}{n_2-2}\sum\limits_{\substack{1\\ i\neq j}}^{N_2} Y_{i2}\,Y_{j2}}{\binom{N_1}{n_1}\binom{N_2}{n_2}}$$

$$+ \frac{\binom{N_1-1}{n_1-1}\binom{N_2-1}{n_2-1}\left(\sum\limits_{i=1}^{N_1} Y_{i1}\right)\left(\sum\limits_{j=1}^{N_2} Y_{j2}\right)}{\binom{N_1}{n_1}\;\binom{N_2}{n_2}}$$

$$= \frac{n_1(n_1-1)}{N_1(N_1-1)} \sum_{\substack{1\\ i\neq j}}^{N_1} Y_{i1}Y_{j1} + \frac{n_2(n_2-1)}{N_2(N_2-1)} \sum_{\substack{1\\ i\neq j}}^{N_2} Y_{i2}Y_{j2}$$

$$+ \frac{n_1}{N_1} \cdot \frac{n_2}{N_2} \Big(\sum_{i=1}^{N_1} Y_{i1}\Big)\Big(\sum_{j=1}^{N_2} Y_{j2}\Big)$$

Then

$$E\Big[\frac{2}{N(N-1)} \Big(\frac{N_1(N_1-1)}{n_1(n_1-1)} \sum_{\substack{1\\ i\neq j}}^{n_1} y_{i1}y_{j1} + \frac{N_2(N_2-1)}{n_2(n_2-1)} \sum_{\substack{1\\ i\neq j}}^{n_2} y_{i2}y_{j2}$$

$$+ \frac{N_1}{n_1} \cdot \frac{N_2}{n_2} \Big(\sum_{i=1}^{n_1} y_{i1}\Big)\Big(\sum_{j=1}^{n_2} y_{j2}\Big)\Big)\Big]$$

$$= \frac{2}{N(N-1)} \Big[\sum_{\substack{1\\ i\neq j}}^{N_1} Y_{i1}Y_{j1} + \sum_{\substack{1\\ i\neq j}}^{N_2} Y_{i2}Y_{j2} + \Big(\sum_{i=1}^{N_1} Y_{i1}\Big)\Big(\sum_{j=1}^{N_2} Y_{j2}\Big)\Big]$$

$$= \frac{2}{N(N-1)} \sum_{\substack{1\\ i\neq j}}^{N} Y_iY_j \tag{4}$$

Therefore,

$$\frac{2}{N(N-1)}\left[\frac{N_1(N_1-1)}{n_1(n_1-1)}\sum_{\substack{1\\ i\neq j}}^{n_1} y_{i1}y_{j1} + \frac{N_2(N_2-1)}{n_2(n_2-1)}\sum_{\substack{1\\ i\neq j}}^{n_2} y_{i2}y_{j2}\right.$$

$$\left.+\frac{N_1}{n_1}\cdot\frac{N_2}{n_2}\left(\sum_{i=1}^{n_1} y_{i1}\right)\left(\sum_{j=1}^{n_2} y_{j2}\right)\right] \qquad (5)$$

is an unbiased estimate of $\frac{2}{N(N-1)}\sum_{\substack{1\\ i\neq j}}^{N} Y_iY_j$, which is the second term of equation (1).

Note that:

$$2\sum_{\substack{1\\ i\neq j}}^{n_1} y_{i1}y_{j1} = 2\sum_{i=1}^{n_1-1} y_{i1}\left(\sum_{j=i+1}^{n_1} y_{j1}\right) = \left(\sum_{i=1}^{n_1} y_{i1}\right)^2 - \sum_{i=1}^{n_1}(y_{i1})^2 \qquad (6)$$

If formula (6) is substituted into (5), and formula (5) and (3) are considered together, we have shown that for collapsing two strata:

$$E\left[\frac{1}{N}\left\{\frac{N_1}{n_1}\sum_{i=1}^{n_1}(y_{i1})^2 + \frac{N_2}{n_2}\sum_{i=1}^{n_2}(y_{i2})^2\right\} - \frac{1}{N(N-1)} \times\right.$$

$$\left\{\frac{N_1}{n_1}\left(\frac{N_1-1}{n_1-1}\right)\left[\left(\sum_{i=1}^{n_1} y_{i1}\right)^2 - \sum_{i=1}^{n_1}(y_{i1})^2\right] + \frac{N_2}{n_2}\left(\frac{N_2-1}{n_2-1}\right) \times\right.$$

$$\left[\left(\sum_{i=1}^{n_2} y_{i2}\right)^2 - \sum_{i=1}^{n_2}(y_{i2})^2\right] + 2\,\frac{N_1}{n_1}\cdot\frac{N_2}{n_2}\left(\sum_{i=1}^{n_1} y_{i1}\right)\left(\sum_{j=1}^{n_2} y_{j2}\right)\Big\}\Big] = S^2 \quad (7)$$

Therefore, an unbiased estimate of S^2, when simple random samples are selected from two strata and we collapse these to form one stratum, is s^2 where

$$s^2 = \frac{1}{N}\left[\frac{N_1}{n_1}\sum_{i=1}^{n_1}(y_{i1})^2 + \frac{N_2}{n_2}\sum_{i=1}^{n_2}(y_{i2})^2\right] - \frac{1}{N(N-1)}\left\{\frac{N_1(N_1-1)}{n_1(n_1-1)}\times\right.$$

$$\left[\left(\sum_{i=1}^{n_1} y_{i1}\right)^2 - \sum_{i=1}^{n_1}(y_{i1})^2\right] + \frac{N_2}{n_2}\left(\frac{N_2-1}{n_2-1}\right)\left[\left(\sum_{i=1}^{n_2} y_{i2}\right)^2 - \sum_{i=1}^{n_2}(y_{i2})^2\right]$$

$$\left. + 2\frac{N_1}{n_1}\cdot\frac{N_2}{n_2}\left(\sum_{i=1}^{n_1} y_{i1}\right)\left(\sum_{j=1}^{n_2} y_{j2}\right)\right\} \quad (8)$$

The following proves that equation (7) of Chapter IV is a consistent estimate of S^2 :

$$\lim_{\substack{n_1\to N_1\\ n_2\to N_2}} (s^2) = \frac{1}{N}\left\{\sum_{i=1}^{N_1} Y_{i1}^2 + \sum_{i=1}^{N_2} Y_{i2}^2\right\} - \frac{1}{N(N-1)}\left\{\left(\sum_{i=1}^{N_1} Y_{i1}\right)^2 - \sum_{i=1}^{N_1} Y_{i1}^2\right.$$

$$\left. + \left(\sum_{i=1}^{N_2} Y_{i2}\right)^2 - \sum_{i=1}^{N_2} Y_{i2}^2 + 2\left(\sum_{i=1}^{N_1} Y_{i1}\right)\left(\sum_{i=1}^{N_2} Y_{i2}\right)\right\}$$

$$= \frac{1}{N(N-1)}\left\{(N-1)\left[\sum_{i=1}^{N_1} Y_{i1}^2 + \sum_{i=1}^{N_2} Y_{i2}^2\right] + \sum_{i=1}^{N_1} Y_{i1}^2 + \sum_{i=1}^{N_2} Y_{i2}^2\right.$$

$$\left. - \left(\sum_{i=1}^{N_1} Y_{i1} + \sum_{i=1}^{N_2} Y_{i2}\right)^2\right\} = \frac{1}{N-1}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\left(\sum_{i=1}^{N} Y_i\right)^2 = S^2 .$$

To prove that equation (11) of Chapter IV is an unbiased estimate of equation (10), the element covariance, consider the following:

$$S_{XY} = \frac{\sum_{i=1}^{N} X_i Y_i - \frac{\left(\sum_{i=1}^{N} X_i\right)\left(\sum_{i=1}^{N} Y_i\right)}{N}}{N-1}$$

$$= \frac{\sum_{i=1}^{N} X_i Y_i}{N-1} - \frac{\sum_{i=1}^{N} X_i Y_i + \sum_{\substack{i \\ i \neq j}}^{N} X_i Y_j}{N(N-1)}$$

$$= \frac{1}{N}\sum_{i=1}^{N} X_i Y_i - \frac{\sum_{\substack{i \\ i \neq j}}^{N} X_i Y_j}{N(N-1)} \qquad (9)$$

$$E\left[\sum_{i=1}^{n} x_i y_i\right] = E\left[\sum_{i=1}^{n_1} x_{i1} y_{i1} + \sum_{i=1}^{n_2} x_{i2} y_{i2}\right]$$

$$= \frac{\binom{N_1-1}{n_1-1}\binom{N_2}{n_2}\sum_{i=1}^{N_1} X_{i1}Y_{i1} + \binom{N_1}{n_1}\binom{N_2-1}{n_2-1}\sum_{i=1}^{N_2} X_{i2}Y_{i2}}{\binom{N_1}{n_1}\binom{N_2}{n_2}}$$

$$= \frac{n_1}{N_1}\sum_{i=1}^{N_1} X_{i1}Y_{i1} + \frac{n_2}{N_2}\sum_{i=1}^{N_2} X_{i2}Y_{i2} .$$

Therefore,

$$E\left[\frac{1}{N}\left(\frac{N_1}{n_1}\sum_{i=1}^{n_1} x_{i1}y_{i1} + \frac{N_2}{n_2}\sum_{i=1}^{n_2} x_{i2}y_{i2}\right)\right]$$

$$= \frac{1}{N}\left\{\sum_{i=1}^{N_1} X_{i1}Y_{i1} + \sum_{i=1}^{N_2} X_{i2}Y_{i2}\right\} = \frac{1}{N}\sum_{i=1}^{N} X_iY_i . \tag{10}$$

An unbiased estimate of the first term of equation (9), when simple random samples are selected from two strata separately which we collapse later to form one stratum, is:

$$\frac{1}{N}\left(\frac{N_1}{n_1}\sum_{i=1}^{n_1} x_{i1}y_{i1} + \frac{N_2}{n_2}\sum_{i=1}^{n_2} x_{i2}y_{i2}\right) . \tag{11}$$

To find an unbiased estimate of the second term in equation (9) consider:

$$E\left[\sum_{\substack{1\\ i\neq j}}^{n} x_i y_j\right] = \frac{\binom{N_1-2}{n_1-2}\binom{N_2}{n_2}\sum\limits_{\substack{1\\ i\neq j}}^{N_1} X_{i1}Y_{j1} + \binom{N_1}{n_1}\binom{N_2-2}{n_2-2}\sum\limits_{\substack{1\\ i\neq j}}^{N_2} X_{i2}Y_{j2}}{\binom{N_1}{n_1}\binom{N_2}{n_2}}$$

$$+ \frac{\binom{N_1-1}{n_1-1}\binom{N_2-1}{n_2-1}\left[\left(\sum\limits_{i=1}^{N_1} X_{i1}\right)\left(\sum\limits_{j=1}^{N_2} Y_{j2}\right) + \left(\sum\limits_{i=1}^{N_1} Y_{i1}\right)\left(\sum\limits_{j=1}^{N_2} X_{j2}\right)\right]}{\binom{N_1}{n_1}\binom{N_2}{n_2}}$$

$$= \frac{n_1(n_1-1)}{N_1(N_1-1)}\sum_{\substack{1\\ i\neq j}}^{N_1} X_{i1}Y_{j1} + \frac{n_2(n_2-1)}{N_2(N_2-1)}\sum_{\substack{1\\ i\neq j}}^{N_2} X_{i2}Y_{j2}$$

$$+ \frac{n_1}{N_1}\cdot\frac{n_2}{N_2}\left[\left(\sum_{i=1}^{N_1} X_{i1}\right)\left(\sum_{j=1}^{N_2} Y_{j2}\right) + \left(\sum_{i=1}^{N_1} Y_{i1}\right)\left(\sum_{j=1}^{N_2} X_{j2}\right)\right]$$

Therefore,

$$E\left\{\frac{1}{N(N-1)}\left[\frac{N_1(N_1-1)}{n_1(n_1-1)}\sum_{\substack{1\\ i\neq j}}^{n_1} x_{i1}y_{j1} + \frac{N_2(N_2-1)}{n_2(n_2-1)}\sum_{\substack{1\\ i\neq j}}^{n_2} x_{i2}y_{j2} + \right.\right.$$

$$+ \frac{N_1}{n_1} \cdot \frac{N_2}{n_2} \left[\left(\sum_{i=1}^{n_1} x_{i1} \right) \left(\sum_{j=1}^{n_2} y_{j2} \right) + \left(\sum_{i=1}^{n_1} y_{i1} \right) \left(\sum_{j=1}^{n_2} x_{j2} \right) \right] \Bigg\}$$

$$= \frac{1}{N(N-1)} \Bigg\{ \sum_{\substack{1 \\ i \neq j}}^{N_1} X_{i1} Y_{j1} + \sum_{\substack{1 \\ i \neq j}}^{N_2} X_{i2} Y_{j2} + \left(\sum_{i=1}^{N_1} X_{i1} \right) \left(\sum_{j=1}^{N_2} Y_{j2} \right)$$

$$+ \left(\sum_{i=1}^{N_1} Y_{i1} \right) \left(\sum_{j=1}^{N_2} X_{j2} \right) \Bigg\} = \frac{1}{N(N-1)} \Bigg\{ \sum_{\substack{1 \\ i \neq j}}^{N} X_i Y_j \Bigg\} \qquad (12)$$

is an unbiased estimate of the second term in equation (9). Notice that

$$\sum_{\substack{1 \\ i \neq j}}^{n_1} x_{i1} y_{j1} = \left(\sum_{i=1}^{n_1} x_{i1} \right) \left(\sum_{i=1}^{n_1} y_{i1} \right) - \sum_{i=1}^{n_1} x_{i1} y_{i1} \qquad (13)$$

If we use identity (13) and the results found in (12) and (11), an unbiased estimate of equation (9), when two strata are being collapsed, is:

$$s_{xy} = \frac{1}{N} \left(\frac{N_1}{n_1} \sum_{i=1}^{n_1} x_{i1} y_{i1} + \frac{N_2}{n_2} \sum_{i=1}^{n_2} x_{i2} y_{i2} \right)$$

$$- \frac{1}{N(N-1)} \Bigg\{ \frac{N_1}{n_1} \left(\frac{N_1 - 1}{n_1 - 1} \right) \left[\left(\sum_{i=1}^{n_1} x_{i1} \right) \left(\sum_{i=1}^{n_1} y_{i1} \right) - \sum_{i=1}^{n_1} x_{i1} y_{i1} \right] +$$

$$+ \ \frac{N_2}{n_2}\left(\frac{N_2-1}{n_2-1}\right)\left[\left(\sum_{i=1}^{n_2} x_{i2}\right)\left(\sum_{i=1}^{n_2} y_{i2}\right) - \sum_{i=1}^{n_2} x_{i2}y_{i2}\right]$$

$$+ \ \frac{N_1}{n_1}\left(\frac{N_2}{n_2}\right)\left[\left(\sum_{i=1}^{n_1} x_{i1}\right)\left(\sum_{j=1}^{n_2} y_{j2}\right) + \left(\sum_{i=1}^{n_1} y_{i1}\right)\left(\sum_{j=1}^{n_2} x_{j2}\right)\right]\Bigg\} \qquad (14)$$

To prove that equation (11) of Chapter IV is a consistent estimate of the element covariance, equation (10), consider the following:

$$\lim_{\substack{n_1 \to N_1 \\ n_2 \to N_2}} \left\{s_{xy}\right\} = \frac{1}{N}\left[\sum_{i=1}^{N_1} X_{i1}Y_{i1} + \sum_{i=1}^{N_2} X_{i2}Y_{i2}\right]$$

$$- \ \frac{1}{N(N-1)}\left\{\left(\sum_{i=1}^{N_1} X_{i1}\right)\left(\sum_{i=1}^{N_1} Y_{i1}\right) - \sum_{i=1}^{N_1} X_{i1}Y_{i1}\right.$$

$$+ \left(\sum_{i=1}^{N_2} X_{i2}\right)\left(\sum_{i=1}^{N_2} Y_{i2}\right) - \sum_{i=1}^{N_2} X_{i2}Y_{i2}$$

$$\left. + \left(\sum_{i=1}^{N_1} X_{i1}\right)\left(\sum_{i=1}^{N_2} Y_{i2}\right) + \left(\sum_{i=1}^{N_1} Y_{i1}\right)\left(\sum_{i=1}^{N_2} X_{i2}\right)\right\}$$

$$= \frac{1}{N(N-1)} \Big\{ (N-1)\Big[\sum_{i=1}^{N_1} X_{i1}Y_{i1} + \sum_{i=1}^{N_2} X_{i2}Y_{i2} \Big] + \sum_{i=1}^{N_1} X_{i1}Y_{i1}$$

$$+ \sum_{i=1}^{N_2} X_{i2}Y_{i2} - \Big(\sum_{i=1}^{N_1} X_{i1} + \sum_{i=1}^{N_2} X_{i2} \Big) \Big(\sum_{i=1}^{N_1} Y_{i1} + \sum_{i=1}^{N_2} Y_{i2} \Big) \Big\}$$

$$= \frac{1}{N-1} \sum_{i=1}^{N} X_i Y_i - \frac{1}{N(N-1)} \Big(\sum_{i=1}^{N} X_i \Big) \Big(\sum_{i=1}^{N} Y_i \Big) = S_{XY} .$$

Therefore, if simple random samples are chosen from two strata separately, an unbiased and consistent estimate of the element covariance of the stratum formed by collapsing the two strata is s_{xy}, defined in equation (14).

BIBLIOGRAPHY

Cochran, William G. Sampling Techniques. Second edition. New York and London: John Wiley & Sons, Inc., 1963.

Hansen, Morris H., Hurwitz, William N., and Madow, William G. Sample Survey Methods and Theory. Vol. I: Methods and Applications. Vol. II: Theory. New York: John Wiley & Sons, Inc.; London: Chapman & Hall, Limited, 1953.

Kish, Leslie. Survey Sampling. New York, London, and Sydney: John Wiley & Sons, Inc., 1965.

Moser, C. A. Survey Methods in Social Investigations. Melbourne, London, and Toronto: William Heinemann Ltd., 1958.

Myers, John G. "Conceptual and Measurement Problems in Job Vacancies: A Progress Report on the NICB Study," The Measurement and Interpretation of Job Vacancies. A Conference Report of the National Bureau of Economic Research. New York and London: Columbia University Press, 1966.

Myers, John G., and Creamer, Daniel. Measuring Job Vacancies. New York: The Conference Board, 1967.

Sukhatme, Pandurang V. Sampling Theory of Surveys with Applications. New Delhi, India: The Indian Society of Agricultural Statistics; Ames, Iowa, U.S.A.: The Iowa State College Press, 1954.

U.S. Bureau of the Census. The Current Population Survey - A Report on Methodology. Technical Paper No. 7. U.S. Government Printing Office, Washington, D.C., 1963.

Wingeard, Irwin F. O. "Experimental Job Vacancy Survey Program of the United States Department of Labor," The Measurement and Interpretation of Job Vacancies. A Conference Report of the National Bureau of Economic Research. New York and London: Columbia University Press, 1966.

Unpublished Material

U.S. Bureau of Employment Security. Findings and Recommendations of the Job Vacancy Experimental Program Conducted During Fiscal Year 1965. Washington, D.C., November, 1965.